Table 1. *Commentators for special sleep and dreams issue*

| | Target article and precommentary authors | | | | | | |
|---|---|---|---|---|---|---|---|
| Commentators | Shepard | Barlow | Hecht | Kubovy & Epstein | Schwartz | Tenenbaum & Griffiths | Todorovič |
| Baddeley, R., Osorio, D., Osorio, D., & Jones, C. D. | | | | | | JBT | |
| Bedford, F. L. | RNS | | | | | | |
| Bertamini, M. | RNS | HB | HH | MK | | | DT |
| Boroditsky, L. & Ramscar, M. | | | | | | JBT | |
| Brill, M. H. | RNS | | | | | | |
| Bruno, N. & Westland, S. | RNS | | | | | | |
| Chater, N., Vitányi, P. M. B. & Stewart, N. | RNS | | | | | JBT | |
| Cheng, K. | RNS | | | | | | |
| Decock, L. & van Brakel, J. | RNS | | | | | | |
| Dowe, D. & Oppy, G. | RNS | HB | | | | JBT | |
| Dresp, B. | RNS | HB | | | | | |
| Edelman, S. | RNS | | | | | | |
| Foster, D. H. | RNS | HB | HH | MK | RS | | DT |
| Frank, T. D., Daffertshofer, A. & Beek, P. J. | RNS | | | | | | |
| Gentner, D. | RNS | | | | | JBT | |
| Gerbino, W. | RNS | | | MK | | | |
| Gold, I. | RNS | | | | | | |
| Heil, J. | RNS | | | | | | |
| Heit, E. | | | | | | JBT | |
| Heschl, A. | RNS | | | | | | |
| Hoffman, W. C. | RNS | HB | HH | MK | RS | | DT |
| Hood, B. | RNS | | HH | | | | |
| Intraub, H. | RNS | | HH | | RS | | DT |
| Jacobs, D. M., Runeson, S. & Andersson, I. E. K. | RNS | | HH | MK | | | |
| Kaiser, M. K. | | | HH | | | | |


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Table 1. *Commentators for special sleep and dreams issue*

| | Target article and precommentary authors | | | | | | |
|---|---|---|---|---|---|---|---|
| Commentators | Shepard | Barlow | Hecht | Kubovy & Epstein | Schwartz | Tenenbaum & Griffiths | Todorovič |
| Krist, H. | RNS | | HH | | | | |
| Kurthen, M. | RNS | | HH | MK | | | |
| Lacquaniti, F. & Zago, M. | RNS | HB | HH | | | | |
| Lee, M. D. | | | | | | JBT | |
| Lomas, D. | RNS | | | | | | |
| Love, B. C. | | | | | | JBT | |
| Massaro, D. W. | RNS | | | | | JBT | |
| Mausfeld, R. | RNS | | | MK | | | |
| Movellan, J. R. & Nelson, J. D. | RNS | | | | | JBT | |
| Niall, K. K. | RNS | | HH | MK | | | DT |
| O'Brien, G. & Opie, J. | | | | MK | | | |
| Pani, J. R. | RNS | | | | | | |
| Pasrons, L. M. | RNS | | | | | | |
| Pickering, J. | RNS | | | | | | |
| Pothos, E. M. | RNS | | | | | JBT | |
| Pribram, K. H. | RNS | | | | | | |
| Raffone, A., Belardinelli, M. O., & Van Leeuwen, C. | RNS | HB | | | | | |
| Schwartz, D. A. | RNS | | | | | | |
| Sokolov, E. N. | RNS | | | | | | |
| Todd, P. M. & Gigerenzer, G. | RNS | HB | | MK | | | |
| Todorovič, D. | | | | MK | | | |
| Vallortigara, G. & Tommasi, L. | RNS | | | | | | |
| Vickers, D. | RNS | | HH | | | | |
| Whitmyer, V. | RNS | | | | | | |
| Wilson, A. & Bingham, G. P. | RNS | | | | | | |
| Wilson, M. | | | HH | MK | | | |
| Zimmer, A. C. | RNS | | | | | | |


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 581–601
*Printed in the United States of America*

# Perceptual-cognitive universals as reflections of the world

**Roger N. Shepard**
*Department of Psychology, Stanford University, Stanford, CA 94305-2130*
**roger@psych.stanford.edu**

**Abstract:** The universality, invariance, and elegance of principles governing the universe may be reflected in principles of the minds that have evolved in that universe – provided that the mental principles are formulated with respect to the abstract spaces appropriate for the representation of biologically significant objects and their properties. (1) *Positions* and *motions* of objects conserve their shapes in the geometrically fullest and simplest way when represented as points and connecting geodesic paths in the six-dimensional manifold jointly determined by the Euclidean group of three-dimensional space and the symmetry group of each object. (2) *Colors* of objects attain constancy when represented as points in a three-dimensional vector space in which each variation in natural illumination is canceled by application of its inverse from the three-dimensional linear group of terrestrial transformations of the invariant solar source. (3) *Kinds* of objects support optimal generalization and categorization when represented, in an evolutionarily-shaped space of possible objects, as connected regions with associated weights determined by Bayesian revision of maximum-entropy priors.

**Keywords:** apparent motion; Bayesian inference; cognition; color constancy; generalization; mental rotation; perception; psychological laws; psychological space; universal laws

## Introduction

The ways in which genes shape an individual's perceptual and cognitive capabilities influence the propagation of those genes in the species' ecological niche just as much as the ways in which those genes shape the individual's physical size, shape, and coloration. A predatory bird has come to have not only sharp talons but also sharp eyes, and a small rodent has come to have not only quick feet but also quick recollection of the location of its burrow. Moreover, natural selection favors adaptation to any biologically relevant property of the world, whether that property holds only within a particular species' local niche or throughout all habitable environments. Thus, both the hawk and the ground squirrel have internalized the period of the terrestrial circadian cycle, whose 24-hour value is the same everywhere on earth and whose invariance is a consequence of a law – the conservation of angular momentum – holding throughout the universe.

From among the general properties that characterize the environments in which organisms with advanced visual and locomotor capabilities are likely to survive and reproduce, here I focus on the following three. (1) Material objects are generally conserved and, when they move (whether relative to the stable environment or to the self-moving observer), move in ways whose possibilities and geometrical simplicities are determined by the three-dimensional, Euclidean character of physical space. (2) The light scattered to an eye from an object's surface bilinearly conflates the invariant spectral reflectance properties of the surface itself and the momentary spectral composition of the illumination, which is subject to three principal degrees of freedom of linear transformation. (3) Objects that are of the same basic kind and, hence, that have the same biologically significant potential (e.g., of being edible, poisonous, predatory, or suited to mating, parenting, and hence propagation of one's genes), generally form a connected local region in the space of possible objects, despite appreciable differences among individual objects of that kind in size, shape, position, motion, or color.

In perceptually advanced mobile organisms, then, genes that have internalized these pervasive and enduring facts about the world should ultimately prevail over genes that leave it to each individual to acquire such facts by trial and possibly fatal error. If so, psychological science may have unnecessarily restricted its scope by implicitly assuming that psychological principles, unlike the universal laws of physics, apply at most to the particular animals that happen to have evolved on one particular planet. When formalized at a sufficient level of abstraction, mental principles that have evolved as adaptations to principles that have long held throughout the universe might be found to partake of some of the generality of those prior principles (Shepard 1987a) – perhaps even attaining the kind of universality, invariance,

ROGER SHEPARD is the Ray Lyman Wilbur Professor of Social Science, Emeritus, of Stanford University and a recipient of the 1995 National Medal of Science. In the early 1960s, he devised the auditory illusion of endlessly rising pitch and developed the first of the methods of "nonmetric" multidimensional scaling, which he has used to reveal hidden structures and functional laws of generalization and perceptual and cognitive representation. Shepard then introduced the paradigm of "mental rotation" and with his students demonstrated the analog character of such imagined transformations. Shepard has sought universal mental principles that reflect universal features of the world.

© 2001 Cambridge University Press 0140-525X/01 $12.50

https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

and formal elegance (if not the quantitative precision) previously accorded only to the laws of physics and mathematics.

My own searches for universal psychological principles for diverse perceptual-cognitive domains have been unified by the idea that invariance can be expected to emerge only when such principles are framed with respect to the appropriate *representational space* for each domain. This idea was inspired, in part, by Einstein's demonstration that in extending physical principles beyond the biologically relevant scales of distance, velocity, mass, and acceleration, invariance could still be achieved – but only by casting those principles in terms of the appropriate four-dimensional space-time manifold. Invariance of the laws of physics was no longer restricted to inertial frames moving at velocities that are small relative to the speed of light (as in Newtonian mechanics, formulated with respect to three-dimensional Euclidean space), or even inertial frames moving at any possible (i.e., subluminal) velocity (as in special relativity, re-formulated with respect to [3+1]-dimensional Minkowsi space). Only when reformulated yet again, with respect to the appropriately curved, [3+1]-dimensional Riemannian space, did the laws of physics finally become (in general relativity) invariant with respect to arbitrarily accelerated frames. Moreover, the motions of objects actually observed in the world were then explained, and explained most accurately (as confirmed, first, in accounting for the perihelion advance of Mercury's orbit and, subsequently, in other ways), not in terms of forces acting instantaneously across arbitrarily large distances in three-dimensional Euclidean space but solely in terms of the local geometry of the curved four-dimensional space-time manifold in the vicinity of the object itself. The paths of motion (like great circles on the surface of the Earth) were now simply the *geodesics*, the direct analogs of straight lines in the curved four-dimensional manifold.

But, for such biologically relevant properties of objects as their positions, motions, shapes, colors, and kinds, what sorts of representational spaces show promise of yielding invariant psychological principles? And if such representational spaces and associated psychological principles arose not accidentally but as adaptations to general properties of the world in which we have evolved, can an identification and analysis of such sources in the world point the way toward elegant and invariant formalizations of the corresponding psychological principles?

## 1. Representations of an object's position, motion, and shape

Position, motion, and shape are best considered together because, from the abstract, geometrical point of view that promises the most elegant and invariant formulation, the representations of these three attributes are inextricably interconnected. I focus initially and most extensively on the representations of positions and rigid motions between positions. Shape I can consider only briefly here, merely observing that the shape of an object may be understood in terms of the object's approximations to all possible symmetries, which in turn may be understood in terms of the object's self-similarities under all possible rigid motions.

The positions, motions, and shapes that are possible for an object depend on the kind of space within which that object is confined. On a biologically relevant scale (of size, velocity, mass, and acceleration), one of the most general facts about the world in which we have evolved is that it is spatially three-dimensional and Euclidean. But how do we demonstrate that humans or other animals have internalized the geometry peculiar to this particular type of space?

The internalization of the circadian cycle was established when animals were raised in artificial isolation from the terrestrially prevailing 24-hour oscillation in illumination and temperature and were found, even so, to maintain a close approximation to their previous 24-hour activity cycle. (As the old quip has it: "You can take the boy out of the country, but you can't take the country out of the boy.") Similarly, the three-dimensionality of our world is so deeply entrenched in our mental makeup that while we may muse, "If only I had a larger office, I would have more room for my books," it does not occur to us to think, "If only I had a four-dimensional office, I would have more degrees of freedom for arranging them!" The very universality of the three-dimensionality of our world precludes our taking "the boy" or, indeed, the girl, the hawk, or the ground squirrel out of this three-dimensional "country," to see whether, in the absence of external support, any of these creatures would continue to perceive and to think three-dimensionally. We can, however, investigate whether an individual, though remaining in three-dimensional space physically, is able to take an object out of that space mentally, when only such a move could achieve compliance with another deeply internalized principle, such as the principle of object conservation.

*Apparent motion*, which is typically induced in an observer by alternately presenting two identically shaped objects in different static positions, provides one means of exploring this possibility. In the absence of any physically presented motion, the particular motion that is experienced must be a direct reflection of the organizing principles of the viewer's brain. The Gestalt psychologists, who were responsible for most of the early studies of apparent motion (see, e.g., Koffka 1931; 1935; Korte 1915; Wertheimer 1912), regarded such organizing principles as manifestations, in the neurophysiological medium of the brain, of minimization principles that operate in physical media generally – much as the spherical shape of a soap bubble arises from principles of conservation of matter (the enclosed volume of air) and minimization of surface area (the enclosing film of soap, with its surface tension). The uniquely powerful organizing principles of the brain are not, however, likely to be wholly explained by properties that grey matter shares with all matter. The neuronal circuits of the brain (unlike the molecules of such media as air or soap films) have been shaped by natural selection specifically to provide a veridical representation of significant objects and events in the external world.

### 1.1. Apparent motion achieves object conservation

Why, for example, does one experience a single object moving back and forth at all, rather than experiencing what is actually being physically presented in the laboratory – namely, two visual stimuli going on and off separately? Quite apart from questions about the particular type of movement experienced, the fact that any connecting movement is experienced is presumably the manifestation of an internalized principle of object conservation. It is simply more probable in our world that an enduring object abruptly moved from one position to a nearby position, than that one object suddenly ceased to exist and, at exactly the


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

same instant, a separate but similar object just as suddenly materialized in another position. Still, if the benefits of representing objects as enduring entities support the instantiation of a connecting motion, two questions remain: Out of the infinity of such possible motions, which particular motion will be instantiated? What formal characterization of that psychologically preferred motion will most elegantly reflect any simplicity, universality, and invariance of its ultimate source in the world?

### 1.2. Apparent motion is experienced in three-dimensional space

When identical two-dimensional shapes, such as the Cooper (1975) polygons adopted for illustration in Figure 1a, are alternately presented in orientationally different positions in their common two-dimensional plane, a single such shape is experienced as rigidly rotating about a fixed point in that plane (e.g., Farrell & Shepard 1981; Robins & Shepard 1977; Shepard 1981b; 1984). Similarly, when identical three-dimensional shapes, such as the Shepard-Metzler (Shepard & Metzler 1971) objects shown in Figure 1c, are alternately presented in their common three-dimensional space, a single such object is experienced as rigidly undergoing a rotational (most generally, a screw-like) motion in that space (Shepard 1984; Shepard & Judd 1976; see also Carlton & Shepard 1990a).

But what happens if the two alternately presented shapes are not identical but *enantiomorphic* – that is, mirror images of each other, like a right and left hand? Asymmetric shapes cannot be transformed into each other by any rigid motion confined to the plane or space in which they reside. They can be brought into congruence there only by a shape-reversing reflection of one of the two objects through some line or plane in their two- and three-dimensional spaces, respectively. Nevertheless, between mirror-image polygons in the plane (Fig. 1b), a rigid motion is still experienced. But it is necessarily experienced as a rotation out of the plane, through the three-dimensional space containing that plane (Shepard 1984). Presumably, we perceptually liberate the object from the two-dimensional plane for two reasons: having evolved in a three-dimensional world, we are just as

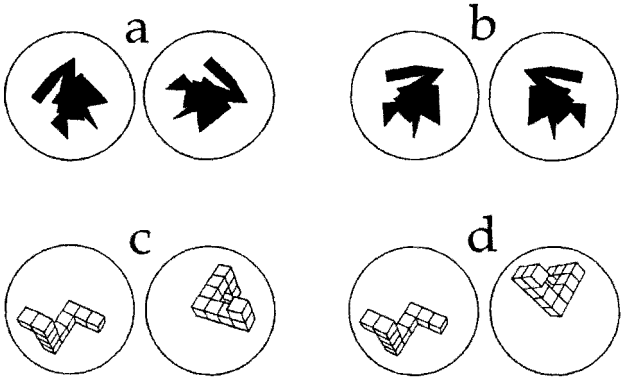



Figure 1. Pairs of alternately presented visual shapes (polygons like those used by Cooper 1975, or block models like those used by Shepard & Metzler 1971) that give rise to four different types of apparent motion: (a) a rigid 90° rotation in the picture plane, (b) a rigid 180° rotation out of the plane and through three-dimensional space, (c) a rigid screw displacement in three-dimensional space, and (d) nonrigid motion only.

capable of representing a rigid motion in three-dimensional space as in a two-dimensional plane. But only the motion in three-dimensional space can represent the shape conservation that is probable in the world – particularly for objects like those in Figure 1, bounded by straight edges or flat surfaces. (This is, incidentally, one reason for our use of stimuli composed of straight lines. The probability that an arbitrarily transformed object will give rise to straight lines in a two-dimensional projection is vanishingly small if nonrigid deformations are allowed. For curved free-form shapes, apparent motion is often experienced as a nonrigid deformation. Moreover, comparison of such shapes by mental rotation is far less accurate – see, for example, Rock et al. 1989.)

Between enantiomorphic solid objects portrayed as in three-dimensional space (Fig. 1d), however, viewers never report experiencing a rigid motion. Such a rigid motion is still mathematically possible – but only by breaking out of the three-dimensional space in which we and our object have been confined, so that we can rigidly rotate the object (now about a plane!) in a surrounding, more commodious four-dimensional space. Failing to achieve even a mental liberation from the only space we have known, we are destined to experience all motions as confined to that three-dimensional space and, hence, all transformations between enantiomorphic shapes as nonrigid. For shapes of the kind illustrated in Figure 1d, at least one of the "arms" of the object typically appears to rotate independently, as if connected to the rest of the object by some sort of swivel joint (a type of motion that, although less common than globally rigid motion, does occur in a world biologically enriched with joint-limbed animals and wind-fractured tree branches).

Similarly, computer generated projections of actual (as opposed to merely apparent) rotations of rigid structures give rise to the "kinetic depth" perception of rigidity for arbitrary rotations in three-dimensional space but not for arbitrary rotations in four-dimensional space (see, e.g., Green 1961; Noll 1965). These phenomena of real and apparent motion (as well as related phenomena of merely imagined motion, e.g., mental rotation) are consonant with the Kantian idea that we are constituted to represent objects and events only in Euclidean space of three (or fewer) dimensions. The modern evolutionary/mechanistic explication of this idea must be that the three-dimensional world simply has not exerted sufficient selective pressures toward the evolution of the more complex neuronal machinery that would be required to represent higher-dimensional spaces and the additional rigid transformations that such spaces afford.

### 1.3. Apparent motion traverses a kinematically simplest path

Even when a connecting motion is possible within three-dimensional space (as in Fig. 1c), the particular motion experienced is only one out of infinitely many possible rigid motions between the two presented positions. One might be tempted to guess that if apparent motion is guided by internalized approximations to principles holding at the biologically relevant scale in the external world, the most likely candidates for those external principles would be those of prerelativistic, Newtonian mechanics. This guess has proved untenable, however, in the face of several facts:

(1) Any rigid motion is compatible with Newton's laws of motion, in the presence of arbitrary unseen forces. Hence, unless we exclude such forces, Newtonian mechanics itself



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

provides no basis for the selection of one path of motion over another.

(2) If we do exclude such forces, however, Newton's laws constrain an object's center of mass to traverse a straight line. But this is contrary to the now well-established finding that apparent motion tends to be over a curved path when the two positions in which the object is alternately presented differ in orientation (see, for example, Bundesen et al. 1983; Farrell 1983; Foster 1975b; Kolers & Pomerantz 1971; McBeath & Shepard 1989; Proffit et al. 1988).

(3) The apparent motions that are most apt to be experienced, as well as the real motions that are discriminated most accurately and judged to be most simple, are those motions whose rotational component is about an axis determined by the geometry of the object's visible shape rather than by the physics of the object's invisible distribution of mass. In particular, the psychologically preferred axes of rotation are those of global or local symmetry of the shape as in Figure 2a – not the principal axes of inertia of the object as in Figure 2c (Carlton & Shepard 1990b). (The latter axes are not even directly determined by the object's visual shape, and can only be inferred by making an additional assumption, such as that the object is of uniform density.) Even an object, such as a cube, for which all possible rotational axes are inertially equivalent appears to rotate about a fixed axis when actually rotated about an axis of symmetry, as in Figure 2b, but appears to wobble when actually rotated about an axis that (though inertially equivalent) is not an axis of geometrical symmetry, as in Figure 2d (Shiffrar & Shepard 1991).

(4) Human infants reveal sensitivity to essentially geometrical constraints such as continuity, rigidity, and impenetrability before manifesting sensitivity to constraints of physical dynamics based on gravity, mass distribution, and inertia (Spelke 1991).

(5) Even adults, from Aristotle to present-day college students, often manifest an "intuitive physics" that fails to comply with the constraints of Newtonian mechanics (McClosky 1983; Proffitt & Gilden 1989; Proffitt et al. 1990; see also Shepard 1987a, pp. 266-67), although in some such cases it may approximate constraints of kinematic geometry (see Shepard 1984; 1987a).

(6) Abstract geometrical constraints apply to a wider range of phenomena in the world than do concrete physical constraints and, for this reason, would presumably have had more opportunity for internalization through natural selection (as well as through learning). Things as lacking in mechanical rigidity as a constellation, a curl of smoke hanging in still air, or a shadow, all undergo transformations that (at least over sufficiently short periods of time) approximate geometrical rigidity relative to a moving or turning observer (Shepard1984; Shepard & Cooper 1982). As Gibson observed, such self-induced geometrical transformations of the "ambient optic array" are probably the most ubiquitous of the transformations with which the visual systems of highly mobile animals must cope (e.g., Gibson 1979). We can understand then why apparent motion might be primarily governed not by the principles of Newtonian mechanics but, rather, by the more abstract and widely manifested constraints of kinematic geometry for three-dimensional space (Shepard 1984).

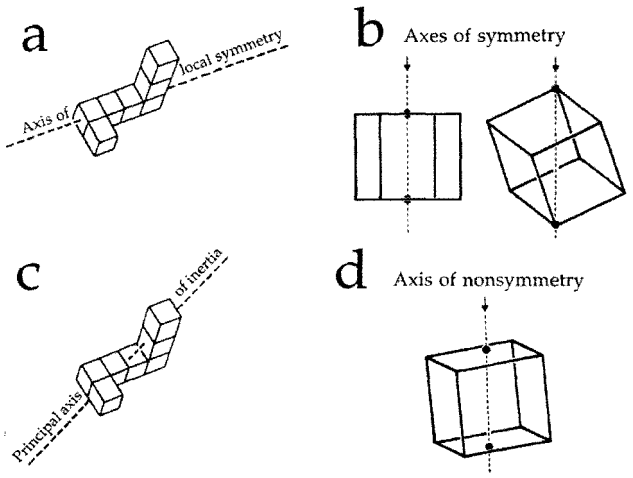



Figure 2. Axes of geometrical symmetry (a) favored by apparent motion and (b) around which real motion appears stable and is accurately compared, and nonsymmetry axes of physical inertia (c) avoided by apparent motion and (d) around which real motion appears to wobble and is less accurately compared. Figures 2a and 2c are from "Psychologically Simple Motions as Geodesic Paths: II. Symmetric Objects," by E. H Carlton and R. N. Shepard, 1990, *Journal of Mathematical Psychology* 34:208. Copyright 1990 by Academic Press. Adapted by permission. Figures 2b and 2d are from "Comparison of Cube Rotations About Axes Inclined Relative to the Environment or to the Cube," by M. Shiffrar and R. N. Shepard, 1990, *Journal of Experimental Psychology: Human Perception and Performance* 7:48. Copyright 1990 by the American Psychological Association. Adapted by Permission.

### *1.4. Kinematic simplicity is determined by geometry*

Kinematic geometry is the branch of mathematics characterizing the motions that are geometrically possible and, among those, the motions that are in a purely geometrical sense most simple or natural – given a geometrical specification both of the object or set of objects and of any constraints on its possible motions. The objects may be geometrically specified to be shape-invariant under all transformations (i.e., rigid). The constraints on their motions may be geometrically specified to preclude mutual interpenetration; escape from their particular embedding space (having specified dimensionality, curvature, and global topology); or violation of the constraints on their relative motions imposed by specified mechanical interconnections (such as a one-degree-of-freedom hinge or slider, a two-degrees-of-freedom pivot, a three-degrees-of-freedom ball and socket joint, etc.). Kinematic geometry says nothing about physical mass, force, acceleration – and, hence, nothing about how much and what kind of effort would be required actually to carry out any particular specified motion, physically, for any given mass distribution within each component object (to say nothing of a specification of the friction at each joint or sliding surface, of the density and viscosity of the medium in which the objects might be immersed, or of how much and what kind of force can be applied before a physical component will bend, fracture, or break). The abstract constraints of geometry are thus conceptually separable from the more concrete constraints of physics: questions of whether a certain large table will fit through a particular door and, if so, what simple sequence of translations and rotations of the table will suffice are purely geometrical and quite distinct from questions of how many persons should be recruited for the job, or of


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

which geometrically possible sequences of rigid transformations will require the least physical effort.

For present purposes, we need consider only the simplest case of the motion of a single rigid object. Even for this simplest case, full mathematical characterization was not achieved until the last century (following the development of the relevant mathematical apparatuses of group theory, Lie algebras, quaternions, and differential geometry). Particularly relevant, here, is Chasles's (1830) theorem of kinematic geometry, according to which any two positions of an asymmetric shape in three-dimensional Euclidean space determine a unique corresponding axis through that space such that the object can be rigidly transported from either position to the other by a combination of a linear translation along that axis and a simple rotation about that same axis – that is, by the helical motion called a *screw displacement*. In particular cases, the translational or the rotational component may be null, leaving only the degenerate screw displacement of (respectively) a pure rotation, a pure translation, or (if both components are null) no motion at all.

If the two positions of an asymmetric object are confined to the Euclidean plane, as in Figure 1a, Chasles's theorem reduces to Euler's theorem. The two positions then determine a unique point in the plane such that the object can be rigidly carried from either position to the other by a simple rigid rotation in the plane about that point. (For generality and elegance, the degenerate case of pure translation is interpreted, in the abstract mathematical formalism, as a rotation of the object about a "point at infinity.")

Strictly, what is uniquely determined by the geometry of the two positions of an (asymmetric) object, is the geodesic path along which a rigid transformation can carry the object back and forth between those positions. Alternative motions along complementary segments of that same geodesic may be possible. Thus, a rotation can carry an object between two positions through either of two nonoverlapping paths around the same circle. Generally, apparent motion tends to be experienced over the shorter of two such alternative paths. But, when the presented positions of the object differ by close to 180°, the two alternatives are of nearly equal length and either motion may be experienced (see Farrell & Shepard 1981; Robins & Shepard 1977). (The case of objects possessing various symmetries, for which two positions of the object may be connected by different screw displacements around two or more distinct axes, will be considered later.) Even when the particular segment of the geodesic over which the motion is to be represented has been determined, kinematic geometry itself does not prescribe the time course of that motion – whether it must be fast or slow, accelerating or decelerating, and so on. In the physical world, the time course of an actual motion is determined by physical dynamics, based on the mass distribution and forces applied. In the mental world, however, the time course of the motion perceptually experienced in apparent motion or only imagined in mental rotation appears to be primarily determined by other, more general, invariant, and adaptively critical constraints, as I shall argue.

Of course, the screw displacements (including simple rotations) prescribed by kinematic geometry are not the only possible motions between two positions of an object in space or in the plane. There are always infinitely many possible motions, including infinitely many rigid motions in which the axes of rotation and translation can vary in orientation from moment to moment and can depart from mutual alignment during the motion, as well as infinitely many more motions that do not preserve the rigid structure of the object. Natural selection has ensured that (under favorable viewing conditions) we generally perceive the transformation that an external object is actually undergoing in the external world, however simple or complex, rigid or nonrigid. Here, however, I am concerned with the default motions that are internally represented under the unfavorable conditions that provide no information about the motion that actually took place between two successive positions of an object. What I am suggesting is that when a simple screw displacement or rigid rotation is possible, that motion will tend to be represented because, of all transformations that conserve the object at the fullest level of shape, it is the geometrically simplest and hence, perhaps, the most quickly and easily computed. Certainly, within a general system suitable for specifying all possible rigid motions, such a motion requires the minimum number of parameters for its complete specification.

### *1.5. Geometry is more deeply internalized than physics*

In accordance with Chasles's theorem, when an asymmetric shape is alternately presented in two orientationally different positions (as in Fig. 1c), under conducive conditions, human viewers generally do report the experience of a helical motion (Shepard 1984). The "conducive conditions" are primarily those in which the temporal interval between the offset of each stimulus and the onset of the other is short enough to yield a pattern of retinal stimulation consistent with some (necessarily rapid) actual motion, and the interval between the onset of each stimulus and the onset of the other is long enough, relative to the extent of the geometrically simplest rigid transformation, to permit completion of the (necessarily rate-limited) neural computations required for that transformation. If the two alternately presented positions of the object are confined to a plane (as in Fig. 1a), the experienced motion generally reduces to a simple rigid rotation around a fixed point in the plane, in accordance with the special case known as Euler's theorem. This single rigid rotation is geometrically simpler than the motion prescribed by Newtonian mechanics, which generally includes two components: a continuous motion of the center of mass (which is rectilinear in the absence of external forces), and an independent rotation about that moving center. Indeed, for a Newtonian motion in three-dimensional space, the axis of rotation need not retain an invariant orientation. Even in the absence of external forces, the axis of momentary rotation will itself wobble about the moving object's center of mass, unless the axis of rotation happens to coincide with a principal axis of inertia of the object. Only in the special case in which the two alternately presented positions of an object have identical orientations does the helical motion prescribed by kinematic geometry coincide with the rectilinear motion prescribed by Newtonian mechanics. Thus, the "intuitive physics" revealed by tests involving spatially extended bodies and rotational motions may deviate from classical physics (e.g., McClosky 1983; Proffitt & Gilden 1989; Proffitt et al. 1990) because whatever internalized knowledge of physical dynamics is tapped by such tests may be contaminated, to a variable degree across individuals and conditions of testing, by a more deeply internalized wisdom about kine-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

matic geometry (Shepard 1984; 1987a; see also Freyd & Jones 1994).

In the method that McBeath and I introduced for quantifying the extent of the departure of apparent motion from a rectilinear path, a shape was alternately presented in different orientations on the left and right of a visual wall and observers adjusted the vertical height of a window in the wall so that the object appeared most compellingly to pass back and forth through that window, which was just large enough to accommodate the object. Figure 3a illustrates the two-dimensional display used in the initial study (see McBeath & Shepard 1989).

The obtained height-of-window settings uniformly implied a curvature away from the straight path, in the direction prescribed by kinematic geometry. As shown in Figure 3b, for linear separations of up to at least 3° of visual angle and for orientational differences of up to at least 90° between the alternately presented stimuli – for which the experience of motion over a particular path was still strong and well-defined – the settings were remarkably close to those prescribed by kinematic geometry. Even for larger separations and angular differences (viz. 180°), for which the experience of motion became weaker and less well-defined, the mean settings remained closer to the circular paths prescribed by Euler's theorem than to the rectilinear paths prescribed by Newtonian dynamics in the absence of external forces. Preliminary indications of similar deviations from rectilinearity have also emerged in subsequent unpublished explorations of the three-dimensional case, where the deviations are generally expected to be helical rather than merely circular. (For example, McBeath, using a computer generated full-color stereoscopic display, had viewers position a circular window anywhere in a two-dimensional wall that appeared to recede in depth, dividing a virtual room into left and right compartments within which the two positions of a three-dimensional object were alternately displayed.)

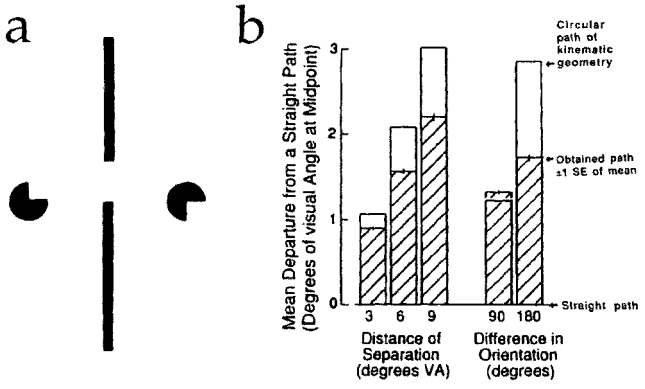



Figure 3. Depictions of (a) shapes alternately presented in different orientations on the left and right of a wall with a window whose height could be adjusted so that a single object appeared to pass back and forth through the window, and (b) mean displacements of the window above the height of a straight path of apparent motion that subjects produced for different linear and angular separations between the shapes. From "Apparent Motion Between Shapes Differing in Location and Orientation: A Window Technique for Estimating Path Curvature," by M. K. McBeath and R. N. Shepard, 1989, *Perception & Psychophysics* 46:334-35. Copyright 1989 by the Psychonomic Society.

### *1.6. Object symmetries entail alternative paths of apparent motion*

For an object possessing some symmetry or symmetries, different screw displacements may be possible between two positions of the object about two or more distinct axes in space. A horizontal rectangular bar in the plane provides a simple illustration. Such a shape is identical to itself under 180° rotation (in the plane) about its center, and under 180° rotations (in space) about either a vertical or a horizontal axis through its center. As a consequence of these symmetries, when such a bar is alternately presented on the left and right, it may be experienced as rigidly moving over any one of seven different paths along five distinct geodesics between the two presented positions, and each of these motions is a screw displacement (if we include, as always, the degenerate screw displacements of pure rotation or pure translation).

Along one geodesic, there are two nonoverlapping 180° rotations in the picture plane around a point midway between the two positions in which the bar appears, one path through the upper portion of the plane, the other through the lower. Along a second geodesic, there are two nonoverlapping 180° rotations in depth about a vertical axis lying in the picture plane midway between the two presented positions, one through the three-dimensional space in front of the plane, the other through the space behind. In each of these first two cases, the two alternative motions correspond to the traversal of two complementary halves of a circular geodesic. Along a third geodesic, two distinct paths of rectilinear translation in the picture plane are geometrically possible between the two positions, one over the short segment of the horizontal line directly between the two side-by-side positions presented, the other over the infinitely longer path corresponding to the complementary part of that horizontal line (interpreted as the complete circle around a "center at infinity"). Finally, along the remaining geodesics, several distinct screw displacements are possible along this same line, in which the bar simultaneously translates and rotates 180° about the short segment of that axis in either direction, or in which the screw displacement entails (again) an infinitely longer translational component over the remaining part of the horizontal line. For these last geodesics, the longer paths of possible transformation, being infinitely longer, are not experienced, leaving just seven likely paths of geodesic transformation.

To obtain experimental evidence that these are the default paths of transformation between two such horizontally separated positions of a rectangular bar, Susan Zare and I primed motions over each of these four paths by appropriately adding a small symmetry-breaking extension to each rectangular bar, giving it the suggestion of one of the four possible L shapes, as shown in Figure 4 (see Carlton & Shepard 1990b, pp. 219-21). With the extension always attached to the upper left corner of the left bar, the symmetry of the right bar could be broken by attaching the corresponding extension to its upper left, upper right, lower right, or lower left (as shown in Fig. 4a, b, c, and d, respectively). The apparent motion tended accordingly to be experienced as a rectilinear translation along the horizontal axis common to the two rectangles (Fig. 4a), as a 180° rotation in depth about the vertical axis lying in the plane halfway between the two rectangles (Fig. 4b), or (less com-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

pellingly, for reasons soon to be noted) as a 180° rotation in the plane about the horizontal line-of-sight axis orthogonal to the plane through a point halfway between the two rectangles (Fig. 4c), or as a 180° screw displacement along the horizontal axis common to the two rectangles (Fig. 4d).

The apparent rotation in the plane (corresponding to Fig. 4c) could also be induced by a form of *path-guided apparent motion* (cf., Shepard & Zare 1983). A low-contrast uniform gray static path was briefly exposed during the 5 msec interval between the offset of each bar and the onset of the other. The path in this case had the shape schematically indicated in Figure 4e by the area that (only for purposes of clear black and white reproduction here) is stippled and much darker than the very light, brief, uniform grey of the path actually presented.

At a random time, while the appropriate induced motion was being experienced, the symmetry-breaking extension (or, alternatively, the faint guiding path) was deleted from the cycling display of the two rectangular bars. Under optimal conditions, viewers typically continued for a few cycles to experience the kinematically simple motion that had been primed by the preceding extensions (or guiding path) before reverting to the experience of either the two most favored default motions, namely, the pure translation indicated in Figure 4a or the pure depth rotation indicated in Figure 4b (see Carlton & Shepard 1990b, p. 220).

The reason that the translation and depth rotation were favored over the rotation in the picture plane (even though both rotations were through the geometrically equivalent 180° angles) is presumably that transformations of the former two types would be more consistent with the retinally available information. For an extended bar, the absence of retinal excitation along any possible connecting motion is less consistent with a rotation in the picture plane (Fig. 4c), for which an actual motion would have tended to stimulate fresh retinal receptors along the path, and more consistent with a translation or a rotation in depth (Figs. 4a and 4b), for which the two presented positions of the bars extensively overlap the path of motion. Even if a fleeting motion had actually occurred over the path corresponding to a rotation in the plane, the resulting weak excitations along the path would have been largely masked by the more forceful retinal "burning-in" of the bar in its more enduring end positions.

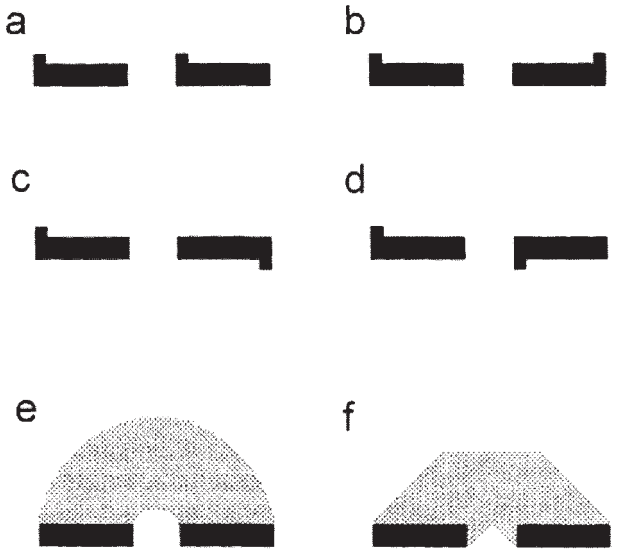



Figure 4. Pairs of alternately presented rectangular bars with L-like extensions that prime four types of apparent motion: (a) rectilinear translation, (b) rotation in depth about a vertical axis, (c) rotation in the picture plane, and (d) a horizontal screw displacement about an axis through the two bars. Pairs of bars with briefly presented interstimulus guiding paths that induce two types of apparent motion: (e) rotation in the picture plane and (f) an up-and-down translation over an inverted V path. (From unpublished experiments by Shepard and Zare.)

### *1.7. Conditions revealing the default paths of mental kinematics*

As I have already noted, natural selection has favored neuronal machinery for swiftly representing whatever motion is actually taking place in the world – not just for representing simple screw displacements. But, to perceive geometrically more complex motions that depart from the default paths of transformation, two conditions must be met: the proximal information must unambiguously specify a more complex distal motion, and the information must impinge on the sensory surface at a rate that does not outstrip the rates of propagation and processing of the neuronal system behind that surface (a system that evolved in a pretechnological world in which most biologically relevant motions were presumably of relatively limited velocity).

Even apparent motion can be induced over a path that does not correspond to a kinematically simple screw displacement. Under appropriate conditions, brief interstimulus presentation of the path schematically illustrated in Figure 4f, for example, can induce a nonrotational experience of the bar translating upward, reversing, and translating back downward in a bouncing inverted-V trajectory between the left and right bar positions. But when the rate of alternation is increased just to the point where the interval between stimulus onsets (the stimulus-onset asynchrony or SOA) becomes too brief for the internal enaction of this kinematically complex motion, the experience tends to revert to the rigid rotation in the plane corresponding to the path depicted in Figure 4e. Presumably, this simple rotation is favored at the shorter SOA because it is the only default motion for which the presented path (Fig. 4f) provides approximate – although not perfect – support. With further reduction of the SOA (or with deletion of the guiding path), the motion usually reverts, once again, either to pure translation or to pure rotation in depth.

From the standpoint advocated by Gibson (1979), apparent motion may seem lacking in ecological validity in a world in which material objects do not go discontinuously in and out of existence. Yet, even in a natural environment, significant objects may be only intermittently visible – as when they are behind wind-blown foliage, for example. One's life can then depend on whether two fleeting visual sensations are interpreted as a single predator moving left to right, or as two distinct objects, one stationed on the left and one stationed on the right. In the laboratory, moreover, the default motions that are experienced in the absence of external support are just the ones that reveal, in their most pristine form, the internalized kinematics of the mind and, hence, provide for the possibility of an invariant psychological law.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

### 1.8. The emergence of invariant laws in representational space

Under appropriate conditions, the minimum time required for representation of a rigid motion between two positions of a stimulus has characteristically increased in an essentially linear manner with the magnitude of the spatial disparity of those positions. Thus, in the case of visual apparent motion, the SOA yielding the experience of a rigid transformation over a connecting path increases approximately linearly with the linear separation between the alternately presented stimuli (e.g., Corbin 1942 [see Shepard 1984, Fig. 5]; Miller & Shepard 1993) or, when the stimuli differ in orientation, with the angular difference between them (Shepard & Judd 1976). Similarly, in the case of mental rotation, the time required to determine whether two objects are identical in shape (as opposed to enantiomorphic) increases approximately linearly with the angular difference in their orientations (see, e.g., Cooper 1975; 1976; Shepard & Metzler 1971). (For overviews of many of the results that have been obtained both for apparent and imagined motion, see, e.g., Cooper & Shepard 1984; Shepard & Cooper 1982; and for an overview of a related phenomenon of "representational momentum," see Freyd 1983.)

Several facts indicate that the slopes of these linear increases of time with distance are not determined by characteristic speeds with which corresponding objects move in the world. There do not seem to be any well-defined characteristic speeds: a bird may perch on a limb or swoop past, a stone may rest on the ground or be hurled. The apparent motion of an object can be experienced before the object itself has been identified as a type likely to move quickly or not at all (e.g., a mouse versus a stone). An object's velocity relative to the observer must, in any case, depend on the observer's own motion. Finally, the obtained slopes of the chronometric functions have generally depended much more on the type of task than on the type of objects presented, with fastest transformational rates found for apparent motion (Shepard & Judd 1976), slower rates for mental rotation (Shepard & Cooper 1982; Shepard & Metzler 1971), and, within mental rotation tasks, slowest rates when two externally presented objects are to be compared (Shepard & Metzler 1971) or when the objects are unfamiliar (Bethell-Fox & Shepard 1988), rather than when an externally presented object is to be compared with an internally represented, already well learned canonical object (Cooper 1975; 1976; Shepard & Metzler 1988).

Again, invariant laws require formulation in terms of more abstract regularities in the world. Neither the path over which an apparent motion is experienced nor the critical time required for the traversal of that path suggests a concrete simulation of the physically or biologically most probable motion of that particular object in that particular circumstance (Shepard 1984). Rather, natural selection seems to have favored the establishment of the identity (or nonidentity) of the two objects in the fastest possible way that preserves whatever is invariant in the structure of the object. Evidently, the fastest possible way for objects in three-dimensional space is via the simplest transformation permitted by the corresponding kinematic geometry of that space. Differences among the rates estimated in the different tasks may not so much reflect differences in typical behaviors of the objects presented as differences in the demands on and external supports for internal computations in those tasks.

The formulation of an invariant chronometric law of linear increase of time with distance requires, of course, that we choose the psychologically appropriate definition of distance. Both for imagined transformation (Shepard & Metzler 1971) and for apparent motion (e.g., Attneave & Block 1973; Corbin 1942; Ogasawara 1936; Shepard & Judd 1976), the appropriate distance evidently is the extent of the relevant transformation in the three-dimensional world more than any distance on the two-dimensional retina. Moreover, invariance is not achieved by defining distance solely in terms of the two objects between which a rigid motion is to be imagined or experienced. Invariance can only be achieved relative to the particular path of motion mentally traversed or experienced on a given occasion, for example, out of all alternative paths that are also permitted by the symmetries of the particular object presented (Farrell & Shepard 1981; Metzler & Shepard 1974 [Fig. 16]; Shepard & Zare 1983).

An invariant chronometric law finally becomes possible when critical times are related to distances along the appropriate geodesic paths in the appropriate representational space. The rate of traversal of such a path is not invariant across different tasks, because natural selection has favored neuronal machinery that yields the fastest possible computation given the external support available, but the external support varies from situation to situation. Even for the same task, the rate is not invariant across different geodesic paths, because no global metric (but only what is called the connection – see Carlton & Shepard 1990a) can be established for the full space of possible positions. (In terms of the formal structure of kinematic geometry, this can be understood by considering that any finite rotation, however small, must dominate any finite translation, however large, because any finite translation is abstractly equivalent to an infinitesimal rotation about a "center at infinity" – Carlton & Shepard 1990a.) For any one given path in the space of possible positions, the linearity of transformation time nevertheless becomes an invariant by virtue of the additive nature of times of analog traversal through successive points along that geodesic. I turn now to a formal characterization of the abstract representational space of possible positions and the geodesics that I take to represent the default paths of apparent or imagined motions. Such a characterization is best developed, first, for the case of an idealized asymmetric object and, then, for the cases of an object's possessing or approximating various symmetries.

### 1.9. The manifold of positions of asymmetric objects, and its geodesics

Objects in three-dimensional space have three degrees of freedom of translation and, except for surfaces of revolution (such as a perfect cylinder, which has an axis of complete rotational symmetry), three additional degrees of freedom of rotation. The complete specification of the position of an asymmetric object at any given moment requires, therefore, the specification of six independent quantities, three for its location and three for its orientation. (Specification of the orientation of a rotationally symmetric ideal cylinder, in contrast, requires only two quantities rather than three, because all angular orientations about its central axis are in-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

distinguishable.) Any rigid motion of an asymmetric object over time thus corresponds to the traversal of a one-dimensional path in an abstract six-dimensional space of the object's distinguishable positions. Moreover, because rotation of any object through 360° returns it to its original position, the three dimensions of orientation are all circular. The abstract six-dimensional space as a whole is accordingly curved and non-Euclidean.

Despite its globally curved, non-Euclidean structure, this six-dimensional space is approximately Euclidean in each local neighborhood – much as the surface of the earth, although globally spherical, approximates a flat Euclidean plane within each sufficiently small region (corresponding, for example, to a single state or country). Spaces that thus approximate Euclidean space in each local neighborhood but may have a globally curved structure are called manifolds. The six-dimensional manifold of object positions has a particular mathematical structure (called, again, its connection) such that the paths in the manifold prescribed by kinematic geometry are the geodesics – the analogs, for a curved space, of straight lines in Euclidean space. (For successive stages in the development of these ideas in connection with the perceptual representation of positions and motions of objects, see Foster 1975b; Shepard 1981b; 1984; Shepard & Farrell 1985; and, most fully, Carlton & Shepard 1990a; 1990b.)

Geodesics are the one-dimensional curves in a manifold that are most simple and uniform in that, like straight lines in Euclidean space, the entire curve can be generated by iteratively applying the same local translational operation that carries any point on the curve into another nearby point on the curve, thus extending the curve in the most natural way. For the geodesics on the surface of a sphere (the *great circles*), for example, a step in the direction that takes one from a point to a nearby point on the geodesic will, with sufficient iteration, take one clear around the circle; equivalently, a straight tape smoothly applied to the surface in the local direction of the curve at any point will eventually return to that starting point, having covered the entire great circle.

As a reflection of the intimate connection between positions and motions that I mentioned at the outset, the set of distinguishable positions of an asymmetric object and the set of rigid displacements of such an object are representable by the same manifold. Once we have selected any one position of an asymmetric object as its canonical reference position, application of any screw displacement (whose rotational component does not exceed 360°) will carry the object into a unique position, and every possible position can be obtained in this way. The correspondence between distinguishable positions and screw displacements is not strictly one-to-one, however. As already remarked, for two objects differing only in orientation, there are two distinct rotations, which will carry one into the other around complementary segments of the geodesic circle. I shall soon return to the consequences of this for the structure of the manifold.

### *1.10. Formal characterization in terms of group theory*

The structure of the set of positions of an object, the set of rigid displacements of the object, and the corresponding manifold with its geodesics can be elegantly formulated in terms of group theory. A group is a set of elements, which in the present case would correspond to rigid displacements of an asymmetric object in space, that meet the following four conditions:

**1.10.1. Closure.** To any ordered pair of elements from the set there is a uniquely corresponding single element, called their *product,* that is also a member of the set. (Thus, for the two screw-displacement transformations, $T_1$ and $T_2$, there is a single such transformation, $T_3$, that carries the object to the same position as the transformation $T_1$ followed by the transformation $T_2$: $T_1 \cdot T_2 = T_3$.)

**10.1.2. Associativity.** An ordered subset of three elements corresponds to the same product element whether a partial product is first formed from the first two elements or from the last two elements, before forming a final product with the remaining element. (Thus, for the ordered set of transformations, $T_1$, $T_2$, and $T_3$: $[T_1 \cdot T_2] \cdot T_3 = T_1 \cdot [T_2 \cdot T_3]$.)

**1.10.3. Existence of identity element.** The set of elements contains a unique element whose product with any given element is just that given element. (Thus the degenerate transformation, here denoted 1, that leaves the position of an object unchanged has no effect beyond the effect of any given transformation, $T_1$, that it precedes or follows: $T_1 \cdot 1 = 1 \cdot T_1 = T_1$.)

**1.10.4. Existence of inverse.** For every element in the set, there is a unique element in the set, called its inverse, such that the product of the element and its inverse is the identity element. (Thus, for every transformation, $T_1$, there is a compensating inverse transformation, $T_1'$, that restores the object to its initial position: $T_1 \cdot T_1' = 1$.)

A familiar example of a group is the set of integers under addition. The group-theoretic "product" in this case is simply the (algebraic) sum of any two integers. Clearly, we have associativity: $(a + b) + c = a + (b + c)$; an identity element (zero); and an inverse for any element n (namely, the integer $-n$). As already implied, the set of elements of a group has dual interpretations – as the set of operations (e.g., the set of continuous displacements in space, or the set of discrete displacements along the number line by addition of positive or negative integers), or as the set of objects obtainable from a canonical element by those operations (e.g., the set of positions of an object in space obtainable by rigid displacements from a reference position, or the set of integers obtainable by integer shifts from – i.e., algebraic additions to – a reference integer, such as zero).

The relevant group for the representation of distinguishable positions or rigid displacements of an asymmetric object in three-dimensional Euclidean space is the Euclidean group, $\mathsf{E}^+$. (The "+" is used here to indicate the restriction to rigid transformations confined within the three-dimensional space, thus excluding reflections between enantiomorphic shapes, such as a left and right hand, that could otherwise be obtained by rigid rotation through a higher-dimensional embedding space.) Because a general screw displacement includes a translational and rotational component, the Euclidean group is composed of the group of linear translations and the group of orthogonal rotations. In group-theoretic terms (see Carlton & Shepard 1990a), $\mathsf{E}^+$ is expressible as the *semidirect product* of the three-dimensional translation group, $\mathbb{R}^3$, and the three-dimensional rotation group $\mathsf{SO}(3)$:



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

$$E^+ = \mathbb{R}^3 \circledS SO(3) \quad (1)$$

The manifold of distinguishable positions (or, equivalently, rigid displacements) of an asymmetric object in three-dimensional space is isomorphic to the Euclidean group, $E^+$.

The concept of the product of two groups may be clarified by considering the simpler product of the group of rigid translations along a line (or, in the discrete case, the group of integers under addition) and the group of rigid rotations about a circle (or, in the discrete case, the group of positive clock-face integers or months of the year 1 through 12, modulo 12). Each of the elements of the *direct product* of these two groups is composed of one element from each of the two component groups (where either element can be the identity element). The direct product of such a rectilinear and circular group is, naturally enough, a cylindrical group. Elements of such a group, by virtue of their rectilinear and circular components, can take us from any point on the surface of the cylinder to any other. In such a direct product group, the elements are *commutative,* that is, the product of two elements is independent of their order so that from a given point on the surface of a cylinder, we get to a given other point whether we first translate the appropriate distance parallel to the axis of the cylinder and then rotate through the appropriate angle about that axis, or whether we first rotate through that angle and then translate over that distance.

In the case of a semidirect product group, however, not all elements will commute in this way. The Euclidean group is necessarily a semidirect product group because rotations in three-dimensional space are generally noncommutative: for an asymmetric shape such as ⫐ (the letter b tipped 90° clockwise in the picture plane), a 90° clockwise rotation followed by a 180° rotation around a horizontal axis yields the result "d" while the same rotations performed in the reverse order – first 180° around a horizontal axis followed by a 90° clockwise rotation in the picture plane – yields the different result "p." (A more complete account of semidirect products is provided by Carlton & Shepard 1990a.)

Each subgroup of a group, such as the subgroup of pure translations $\mathbb{R}^3$ and the subgroup of pure rotations $SO(3)$ of the Euclidean group $E^+$, individually satisfies the already-stated conditions for a group. The Euclidean group also contains other, more restricted subgroups such as the group of translations along a horizontal axis of three-dimensional space, or the group of rotations about a vertical axis of that space.

Of greatest relevance, here, is the set of *one-parameter subgroups* of the Euclidean group. These correspond to the geodesics in the manifold of distinguishable positions, and are straight lines in the three-dimensional translation subgroup, $\mathbb{R}^3$, circles in the three-dimensional rotation subgroup $SO(3)$ and, more generally, helical curves in the full six-dimensional Euclidean group $E^+$. (The designation of these subgroups as "one-parameter" corresponds to the fact that a single parameter suffices to specify a location along a one-dimensional geodesic.) In an analogous but more easily imagined, lower-dimensional, and direct-product case, a tape started at an arbitrary angle will wind helically around the surface of a cylinder, which also has a straight (axial) component, the analog of $\mathbb{R}^3$, and circular (angular) component, the analog of $SO(3)$.

For pure rotations of an object in space, we need consider only the great-circle geodesics in the three-dimensional submanifold corresponding to $SO(3)$. Figure 5 illustrates, by means of the orientations of a labeled cube, a two-dimensional section through this submanifold. The portrayal of this submanifold as a flat disk is only for convenience of illustration in a flat picture. The intrinsic metric of this two-dimensional submanifold is actually that of a spherical surface, thus providing for the great-circle shapes of the geodesics (see Carlton & Shepard 1990a). Moreover, diametrically opposite points around the perimeter of the disk correspond to the same orientation of the object (as shown in the figure by agreement in orientations of the letter B on the back of the cube) and such pairs of points, although widely separated in the figure, should be regarded as the same points.

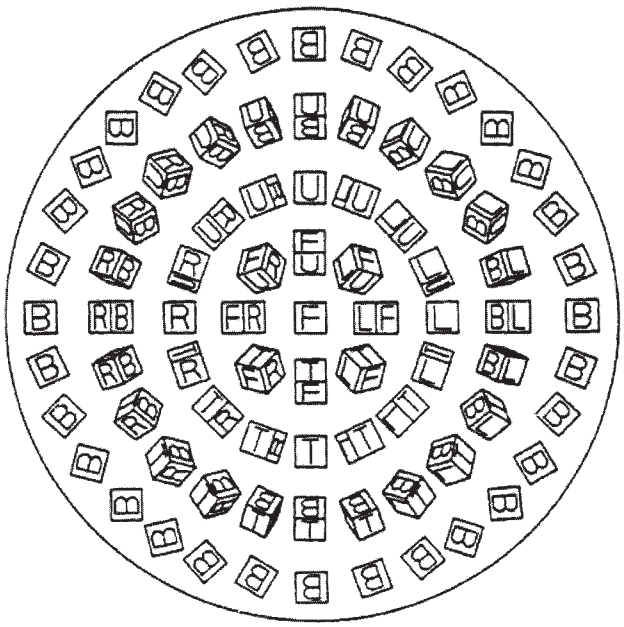


Figure 5. Flattened depiction of one two-dimensional section through the three-dimensional manifold, SO(3), of orientations of a marked cube. From "Representation of the Orientations of Shapes," by R. N. Shepard & J. E. Farrell, 1985, *Acta Psychologica* 59:109. Copyright 1985 by Elsevier Science Publishers. Reproduced by permission.

We are now in a position to clarify further the relation between the spatial representation of distinguishable positions of an asymmetric object and the representation of its rigid displacements. The hemispherical surface illustrated, in flattened form, in Figure 5 includes points corresponding to rotations of only up to 180° from the orientation of the cube represented by the central point (with the F-marked face upright and in front) taken as its canonical orientation. This is sufficient for the representation of all distinguishable orientations falling on geodesics in this surface because, for every rotation through more than 180° (the longer way around a geodesic circle in this surface), there is a rotation through less than 180° (the shorter way around that circle) that is included in the surface and that results in exactly the same orientation of the object. So, although the two possible transformations (the longer and shorter ways around the circle) are distinct, the results of these two transformations are identical. For the complete representation of the three-dimensional subgroup of distinct rota-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tions, SO(3), then, each two-dimensional hemispherical section, such as that illustrated in Figure 5, must have its missing half added, to form a complete sphere. In the complete manifold of rotations, then, diametrically opposite points correspond to distinct rotations (the shorter or longer ways around the same geodesic circle) but, in the corresponding manifold of distinguishable positions, such diametrically opposite points, because they correspond to indistinguishable orientations, are identified (treated as the same point). (Counterintuitively for us, who have evolved to deal with macroscopic objects, such an identification is not needed for an important class of microscopic objects, viz., fermions, which include such basic constituents of matter as electrons and protons. As was first called to my attention by Eddie Oshins, according to an empirically verified prediction of quantum mechanics, these particles do not become physically identical to themselves until rotated through two complete 360° turns!)

### 1.11. Formal characterization of positions and motions for a symmetric object

An object possessing one or more symmetries entails a modification of the manifold of distinguishable positions. By definition, whereas an asymmetric object becomes distinguishable from itself under a rotation through any nonzero angle short of 360°, a symmetric object becomes identical to itself under some other rigid transformation, such as a 180° rotation in the case of a rectangle. Consequently, some widely separated points of the manifold of distinguishable orientations for an asymmetric object (such as the points corresponding to 180° – different orientations of the asymmetric polygon in Fig. 6a) must be mapped onto the same point of the manifold of distinguishable orientations of a symmetric object (such as the single point corresponding to any two 180° – different orientations of the centrally-symmetric polygon in Fig. 6c). As illustrated at the bottom of the figure, the great circle corresponding to one complete picture-plane rotation of the asymmetric polygon (Fig. 6a) is thus twisted (through the intermediate curve shown in Fig. 6b) into a double-wound circle (Fig. 6c) in which each pair of orientations of the polygon separated by 180° maps into the same point (Shepard 1981b; Shepard & Farrell 1985). One complete 360° rotation of a centrally symmetric object (like the polygon in Fig. 6c) is thus represented by two complete excursions around a geodesic circle in the space of distinguishable positions of that object (the circle depicted at the bottom of Fig. 6c).

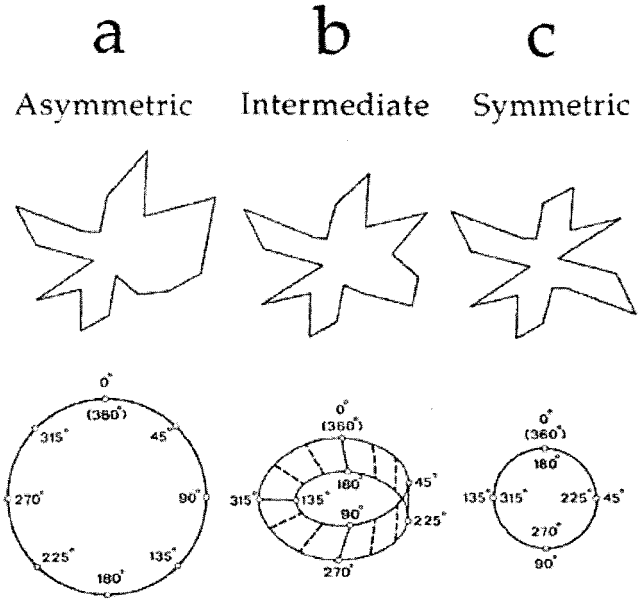



Figure 6. Illustrative polygons (above) and their corresponding geodesic paths of rotation in the picture plane (below) for three degrees of approximation to central symmetry: (a) 0%, (b) 75%, and (c) 100%. From "Psychophysical Complementarity," by R. N. Shepard, 1981b. In: M. Kubovy & J. Pomerantz (Eds.), *Perceptual Organization*, p. 317. Hillsdale, NJ: Lawrence Erlbaum. Copyright 1981 by Lawrence Erlbaum Associates. Adapted by permission.

For an object possessing a symmetry, the submanifold of orientations is necessarily replaced by a *quotient manifold.* Designating these manifolds by the names of their corresponding groups, we can more specifically say that the manifold

$$\mathrm{SO(3)} \text{ is replaced by } \mathrm{SO(3)/S(O)} \qquad (2)$$

where S(O) is the manifold corresponding to the symmetry group of the object (see Carlton & Shepard 1990b). The symmetry group of the object is, simply, the subgroup of rigid transformations that leaves the object indistinguishable from its initial state. Thus the symmetry group of the square is a subgroup of the Euclidean group that includes rotations through 90° and 180° in the plane, as well as 180° rotations in space about vertical, horizontal, and diagonal axes of the square.

Quantitative evidence from a number of experiments (including experiments on real and merely imagined motion, as well as experiments on visual apparent motion) now indicates that psychologically preferred paths of rigid transformation do correspond to geodesics in the appropriate manifold – including the appropriate quotient manifold SO(3)/S(O) for objects with various symmetries (e.g., Farrell & Shepard 1981; Shepard 1981b; Shiffrar & Shepard 1991; see also Carlton & Shepard 1990b, pp. 219-21). As will be noted, such manifolds and geodesics can even be recovered by applying methods of multidimensional scaling to psychological data.

### 1.12. Formal characterization for approximations to various symmetries

Most of the objects that we encounter in the world are neither completely asymmetric nor exactly symmetric. Instead, they more or less approximate various global or merely local symmetries. Just as a strict symmetry of an object corresponds to the transformation (rotation or reflection) that carries that object exactly into itself, a symmetry that is only approximate corresponds to the transformation that achieves a local maximum of correlation in shape between the object and itself – with the degree of approximation measured by the magnitude of the correlation at that local maximum. Only a perfect sphere is identical to itself under every rotation and reflection about its center and, hence, is wholly symmetric. (Thus there is a more abstract, purely geometrical basis of the spherical shape of the soap bubble invoked by the Gestalt psychologists.) A person's face, body, and brain only approximate but do not achieve strict bilateral symmetry. Complete asymmetry, on the other hand, can



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

never be attained. Any shape (including the "random" polygon in Fig. 6a) necessarily resembles itself to greater or lesser degrees under various angles of relative rotation. (Hence, the perfect circle depicted below that polygon does not precisely correspond to the internal representation of the set of distinguishable orientations of that particular polygon. Strictly, that circle is a kind of average representation of possible orientations for a total ensemble of polygons generated by the same random rules.)

Indeed, the shape of any particular object can be defined in terms of its degrees of approximation to all possible symmetries of that object, via the correlation between the object and itself under each possible rotation and reflection (Shepard 1981b; 1988). Although degrees of approximation to symmetries thus appear to be fundamental in human perception and cognition, the classical development of the group-theoretic basis of symmetry in mathematics has treated each type of symmetry as a discrete feature that an object possesses either wholly or not at all.

A formal, quantitative treatment of approximations to symmetries can, however, be given in terms of representational space (Carlton & Shepard 1990b; Farrell & Shepard 1981; Shepard 1981b; 1988; Shepard & Farrell 1985). Approximations to symmetries are regarded as inducing deformations in the original manifold of distinguishable orientations corresponding to SO(3), for an ideally asymmetric object (or ensemble of random objects), toward the manifold, corresponding to SO(3)/SO, for each type of symmetry that a given object approximates.

Farrell and I sought empirical support for such a spatial representation of the orientations of polygons possessing various degrees of approximation to central symmetry, that is, to self-identity under 180° rotation in the two-dimensional plane. The minimum SOAs for the experience of rigid rotational motion between two alternately presented orientations (Farrell & Shepard 1981) and the times required for the discrimination of sameness or difference of two simultaneously presented orientations (Shepard & Farrell 1985) were both consistent with the representations of these shapes in their corresponding manifold of distinguishable positions (see Carlton & Shepard 1990b; Shepard 1981b). Specifically, multidimensional scaling of the discrimination times (using the INDSCAL method of Carroll & Chang 1970) yielded points in four-dimensional space falling close to the particular geodesics prescribed (Shepard & Farrell 1985), namely, closed curves forming the edge of a one-sided Möbius band (as illustrated in two-dimensional projection at the bottom of Fig. 6b); and the minimum SOAs for rigid apparent motion were as predicted for motions between the two alternately presented orientations over just these geodesic paths (Farrell & Shepard 1981).

#### *1.13. Formal connections between the representations of positions, motions, and shapes*

That the same manifold can represent both the distinguishable positions of an object in space and the possible rigid displacements of the object between its distinguishable positions holds also for objects that approximate various symmetries. This is the basis of the inextricable connection noted between the representations of the positions and kinematically simplest motions of an object. Shapes, however, can have many more than six degrees of freedom. Clearly then, shapes cannot be fully represented as individual points in the manifold of positions/displacements, a manifold that has no more than six dimensions (and fewer, for objects, such as a cylinder or a sphere, with complete rotational symmetry). An isomorphism does, nevertheless, hold between the shape of any object and the conformation of the corresponding manifold of positions/displacements for that object. The conformation is dictated by the object's degrees of approximation to all possible symmetries (Shepard 1981b; 1988).

Ultimately, shapes themselves should be formally representable as points in a higher-dimensional manifold of all possible shapes. The full development of such a representation must provide for a detailed, parametric characterization of the degrees of approximation of a shape to possible symmetries in three-dimensional Euclidean space. Just as the position of any given object can be represented, historically, as the result of the simplest rigid transformation that might have carried the object into its given position from a prespecified canonical position, the shape of any given object might be interpreted, historically, as the result of the simplest nonrigid deformation that might have brought the object into its present shape from some prespecified, simplest canonical shape.

Leyton (1992) has achieved significant progress toward a group-theoretic account of how objects may be perceived and represented in terms of the derivational history that each implies. In the spirit of the approach I have outlined here, the appropriate representational space might provide, in general, for the interpretation of any object as having emerged from some more symmetrical, canonical progenitor through the traversal of a symmetry-breaking geodesic in that space. Unlike the manifold of positions and rigid motions, the space of possible shapes and nonrigid motions would be not only higher-dimensional but also anistotropic and inhomogeneous. In a possible, if remote analogy with general relativity, for which a test particle follows a geodesic toward a gravitational singularity in the space-time manifold, the cognitive interpretation of a given shape might be regarded as following a geodesic backward toward a point of maximum symmetry (and, perhaps, minimum entropy) in the manifold of possible shapes.

### 2. Representation of an object's color

The problem of color has inspired major efforts by some of the greatest scientists of all times, including Newton, Young, Helmholtz, Maxwell, and Schrödinger, to name just a few of the most illustrious physicists. So much attention to color might seem difficult to justify from an evolutionary standpoint. The perception and representation of the positions and motions of objects in space is clearly essential for our survival and reproduction. But the perception and representation of colors, though doubtless contributing to our discrimination of some biologically relevant objects (such as red berries against green leaves) and our recognition of, or learned attachment to, others (such as a face with blue, green, or brown eyes, or surrounded by yellow, red, brown, or black hair), evidently is not essential for many animals, including humans.

Originally, the investigation of color was probably motivated, instead, by the challenge of reconciling the seemingly unanalyzable subjective experience of colors with


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

such seemingly colorless concepts of physical science as space, time, particles, or waves – including the electrochemical events in what has aptly been styled "the dark chamber of the skull" (as by B. P. Browne, quoted in William James, 1890 [p. 220 of 1950 edition]). The challenge remains (Shepard 1993), and is even augmented by the need to provide an evolutionary explanation for the ways in which the internal representation of colors differs from the physical characteristics of external surfaces and of the electromagnetic radiations that they reflect in the world.

In this regard, two facts about the human perception of an object's color are perhaps most fundamental: first, the color appearance of an object's surface is essentially invariant despite enormous variations in the spectral composition of the light that falls on that surface and, hence, of the light that the surface scatters back to our eyes. Second, although these physical variations are also potentially of high dimensionality, we can match the color appearance of any such surface by adjusting just three chromatic components produced by a suitable color mixing apparatus. The color appearances of surfaces thus correspond to relatively fixed points in a three-dimensional *color space.*

Schematically, this color space can be thought of as approximating the idealized spherical solid portrayed in Figure 7. We can describe this space either in terms of three cylindrical coordinates of lightness, hue, and saturation (as shown in Fig. 7a), or in terms of three rectangular coordinates of lightness-versus-darkness, redness-versus-greeness, and blueness-versus-yellowness (as indicated in Fig. 7b). But *what in the world* is the source of the three-dimensionality of this color representation? And *what in the world* is the source, in this representation, of the circularity, discovered by Newton, in the continuum of hues? For, this circularity presents us with the psychophysical puzzle that the hues corresponding to the most widely separated of the visible physical wavelengths, namely red and violet, appear more similar to each other than they do to a hue of intermediate wavelength, such as green.)

As before, the answers that may first spring to mind may not necessarily be the correct ones. In the case of motion, the most deeply internalized constraints evidently are determined less by Newtonian mechanics and the mass distribution of each object than by the more abstract kinematic geometry of three-dimensional Euclidean space and the symmetry groups of objects. Similarly in the case of color, I suggest that the three-dimensionality and circular structure of the representation derives less from anything in the intrinsic spectral characteristics of surfaces or of their reflected light than from the more abstract constraints of the universally linear way in which illumination from an invariant stellar source is transformed by a planetary environment, and by the prevailing three-dimensional structure of the planetary transformations. I begin with a consideration of the universal linearity of spectral transformation and the selective pressure toward its internal representation in diurnal animals with highly developed visual systems.

### 2.1. Formal characterization of the linearity underlying color constancy

The invariant physical characteristic of a surface underlying its perceived color is its spectral reflectance function $S^x(\lambda)$. This function specifies for each wavelength of incident light the fraction of that light that will be scattered back to any receptive eyes. Accordingly, the amount of light reaching an eye from a point $x$ on an environmental surface, $P^x(\lambda)$, is expressible as the product of the amount of light of that wavelength in the ambient illumination, $E(\lambda)$, and the spectral reflectance of the surface for that wavelength at point $x$, $S^x(\lambda)$:

$$P^x(\lambda) = E(\lambda)\, S^x(\lambda) \tag{3}$$

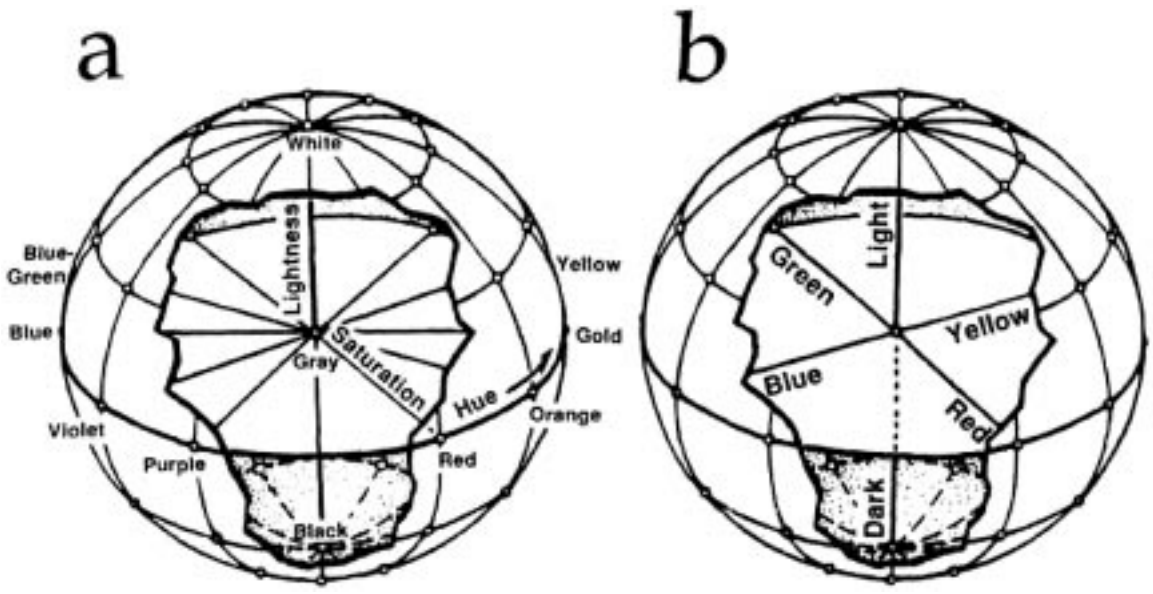



Figure 7. Schematic illustrations of human color space showing (a) its three cylindrical dimensions of lightness, saturation, and hue, and (b) its three opponent-process rectangular dimensions of light-dark, red-green, and blue-yellow. From "The perceptual organization of colors: An adaptation to the regularities of the terrestrial world?" by R. N. Shepard, 1992. In: J. Barkow, L. Cosmides, and J. Tooby (Eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture,* pp. 467–97. New York: Oxford University Press. Copyright 1992 by Oxford University Press. Adapted by permission.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Empirical data and theoretical considerations (concerning universal quantum mechanical interactions between photons and surface molecules), reviewed by Maloney (1986), indicate that the spectral reflectance functions $S_x$ of wavelength for natural surfaces can be approximated as linear combinations of a small number, *n*, of *reflectance basis functions:*

$$S^x(\lambda) = \sum_{j=1}^{n} \sigma_j^x S_j(\lambda), \tag{4a}$$

where $\sigma_j^x$ is the weight for the *j*th of the *n* basis reflectance functions for surface point *x* (see also Brill 1978; Buchsbaum 1980; Sälström 1973).

Other empirical data and theoretical arguments (to which I shall return) indicate that the spectral distributions E for natural conditions of illumination can similarly be approximated as linear combinations of a small number, *m*, of *lighting basis functions:*

$$E(\lambda) = \sum_{i=1}^{m} \epsilon_i E_i(\lambda), \tag{4b}$$

where $\epsilon_i$ is the weight for the *i*th of the *m* basis lighting functions for the ambient illumination (see Maloney & Wandell 1986).

Substitution of Equations 4a and 4b into Equation 3 then yields a dimensionally reduced linear model governing the way illumination and surface properties are combined in the proximal stimulus $P_x$. Figure 8 is my schematic illustra-

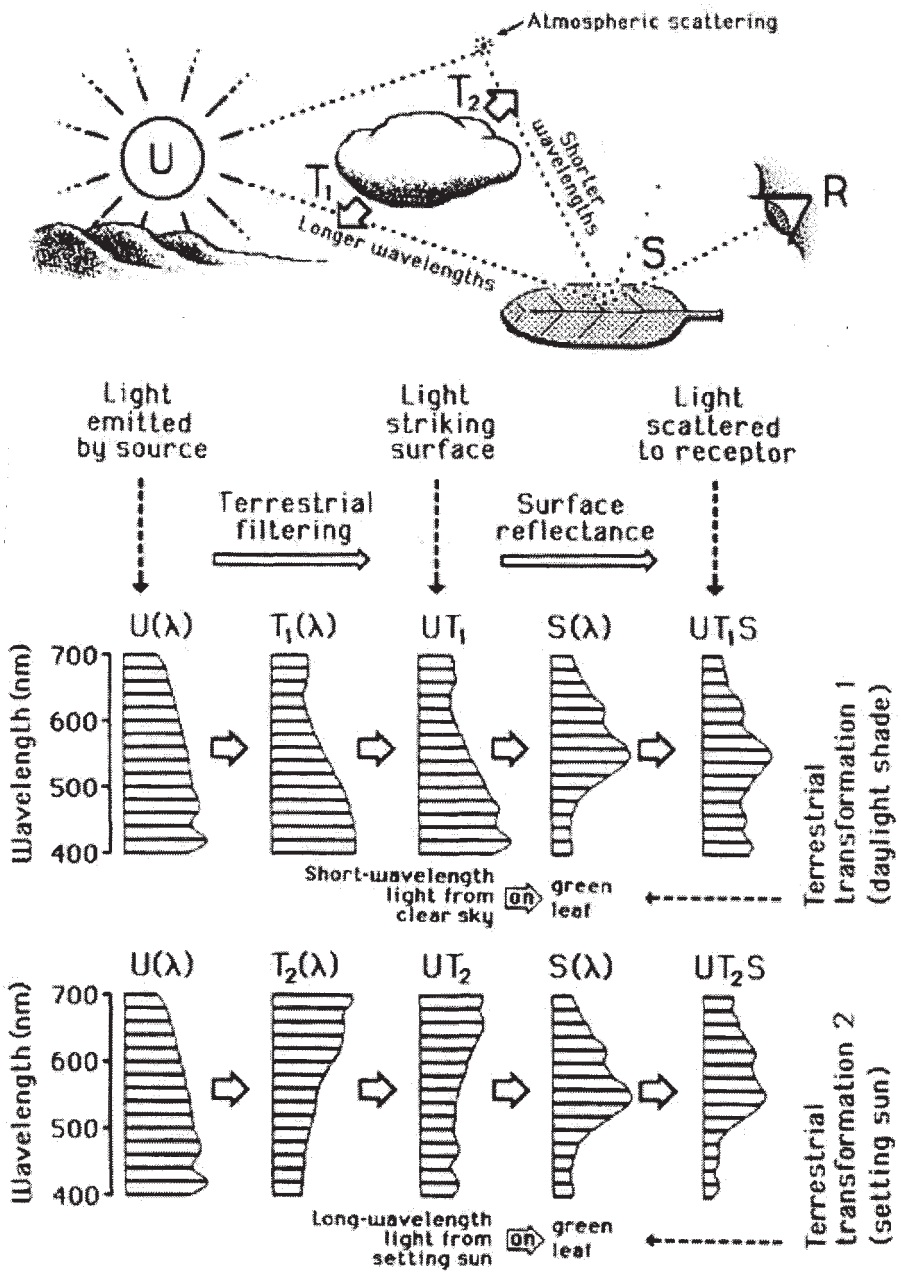



Figure 8. Illustration, for two conditions of terrestrial filtering (T1 and T2), of how the spectral composition of the light from an unvarying source, U, is linearly transformed first by terrestrial filtering, T, and then by scattering from a surface, S, before reaching the observing receptor, R.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tion of how the spectral composition of the light scattered to the eye from a surface (here, a green leaf) differs between two conditions of terrestrial filtering of the illumination, in which a cloud moves to block the longer-wavelength (redder) rays direct from a setting sun (T1), or to block the shorter-wavelength (bluer) rays scattered from the molecules of air (T2). In either case, the distribution of the unvarying solar light (U) is linearly transformed (UT) by the spectral distribution of the momentary terrestrial condition of filtering (T1 or T2), and that light is then linearly transformed again (UTS) by the spectral reflectance distribution of the surface (S) in the process of being scattered to the observing receptor (R). In order to achieve color constancy, the visual system must arrive at a correct characterization of the invariant spectral reflectance distribution $S(\lambda)$ of the external surface (of the leaf) despite the contamination of the spectral distribution of the proximal stimulus (UTS) by the terrestrial filtering $T(\lambda)$ of the illumination.

Using such a linear model, Maloney and Wandell (1986) showed how (under quite general conditions) a visual system with a sufficient number of chromatically distinct types of photoreceptors (such as the red-, green-, and blue-sensitive cones in the human retina) can achieve a disentanglement of the characteristics of the surface (S) and the characteristics of the illumination (UT) and thus attain color constancy. Because the linearity of the way in which the spectral properties of illumination and surface combine is universal, this linearity should tend to be internalized in the visual systems of organisms wherever natural selection has favored color vision. But we are still left with the question of what it is in the world that determines the dimensionality of color representation.

### 2.2. The representation of surface colors should have the dimensionality of natural illumination

According to Maloney (1986), the number of degrees of freedom of spectral reflectance of natural surfaces ($n$ in Eq. 4a) falls somewhere between five and eight. A visual system that completely recovers the chromatic characteristics of such surfaces by means of the computation described by Maloney and Wandell (1986) would require a number of chromatically distinct types of photoreceptors that is one greater than this number of degrees of freedom, that is, between six and nine. The conclusion is clear: a visual system, like ours, that has only three types of color receptors (the red, green, and blue cones) and, hence, that is restricted to three dimensions of color representation cannot fully capture the intrinsic reflectance properties of natural surfaces.

Suppose, however, that the dimensionality of color representation has been favored not because it captures the full spectral reflectance properties of natural surfaces but because it is the minimum dimensionality needed to compensate for natural variations in illumination and, thus, to achieve *constancy* of whatever chromatic aspects of the surfaces are represented. Then, even though we may not perceive everything that could be perceived about each surface, we at least perceive each surface as the same under all naturally occurring conditions of illumination.

Available evidence indicates that the number of degrees of freedom of terrestrial lighting ($m$ in Eq. 4b) is essentially three. Principal components analyses have revealed that the great variety of spectral energy distributions of natural illumination measured for different atmospheric conditions and times of day can be well approximated as linear combinations of just three basic functions (see Judd et al. 1964; and other studies cited in Maloney & Wandell 1986, Note 17). Moreover, the three dimensions of spectral variation in natural illumination have identifiable sources in the world (Shepard 1992):

(1) There is an overall light-versus-dark variation ranging from the full and direct illumination by midday sun and unobstructed sky to whatever portion of that same illumination (uniformly reduced across all wavelengths) reaches an object only by scattering from achromatic clouds, cliffs, or moon.

(2) There is a red-versus-green variation depending on the balance between the long (red) wavelengths, which are selectively passed by atmospherically suspended particles (a particularly significant factor when the sun is close to the horizon) or which are selectively blocked by water vapor, and the remaining band of visible wavelengths, which (ranging from yellows through blues) center on green.

(3) There is a blue-versus-yellow variation depending on the balance between the short (blue and violet) wavelengths, which are selectively scattered (e.g., to a shaded object) by the molecules of the air itself, and the remaining band of visible wavelengths of light directly from the sun (as might reach an object through a small "window" in a leafy canopy), which (ranging from greens through reds) center on yellow.

Possibly, then, the light-dark, red-green, and blue-yellow *opponent processes,* proposed by Hering (1887/1964) and Hurvich and Jameson (1957), on quite different (psychophysical and, subsequently, neurophysiological) grounds, may not have to be accepted as an arbitrary design feature of the human visual system. Such a three-dimensional representation of color may have emerged as an adaptation to a pervasive and enduring feature of the world in which we have evolved. At the same time, this feature may be the nonarbitrary source of the transformation of the rectilinear continuum of physical wavelength into Newton's circle of hues: Very schematically, the two colors in either the red-green pair or the blue-yellow pair, corresponding to the two extremes of variation on an independent dimension of terrestrial filtering, are analogous to diagonally opposite corners of a square (see Shepard & Carroll 1966 [Fig. 6, p. 575]) or diametrically opposite points on a circle (as in Fig. 11 in the next section of the present article). As such, the two opposite colors in either of these pairs must be further apart than the colors in any other pair, including red and blue, which, although corresponding to points close to the extreme ends of the physical continuum of visible wavelengths, are perceptually represented in a way that is more analogous to points separated by one edge of a square or by only a quarter of a circle (Shepard 1992; 1993).

If the linear transformations of the illumination that occur in nature have just three degrees of freedom, then three dimensions are required to compensate for those transformations and, thus, to maintain constancy in the apparent colors. Indeed, three such dimensions are needed to maintain constancy even in just the apparent lightnesses of surfaces, without regard to chromatic color (Shepard 1990; 1992). That is, for every terrestrially induced linear transformation on the illumination, a compensating (inverse) transformation must be internally performed to achieve invariance of the final internal representation of the colors – including even their ordering with respect to achromatic



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

lightness. (Even in the purely achromatic, i.e., grey-scale, representation of, say, a red and a blue surface, the red surface would appear a lighter grey than the blue surface in the light of the setting sun but the blue surface would appear lighter grey than red in the light scattered only from the sky – unless the input is first analyzed into separate chromatic channels and appropriately transformed, before being reduced to the final grey-scale representation.)

Figure 9a indicates how one possible linear transformation (for simplicity of illustration, a two-dimensional transformation representable by a diagonal matrix) would affect the amounts of light of long and short wavelengths reflected back from each of a number of colored surfaces (indicated by the dots). These amounts might correspond to what would be picked up by red and blue cones in the human retina. Under a shift in natural illumination in which, for example, clouds that were blocking light directly from a low sun shift to block, instead, light scattered from the sky, the light scattered back to the eye from each surface (indicated by a filled circle) contains a reduced portion of its original short-wavelength (blue) component and increased portion of its original long-wavelength (red) component (as indicated by the arrow carrying the filled circle to the position of a corresponding open circle). An inverse linear transformation (mapping the rectangle with dashed outline back to the square with solid outline) will reinstate the original configuration of dots and, hence, achieve constancy of appearance of the surface colors. (The chromatic information about the surfaces is not in any sense used up in correcting for the illumination. Because only a small subset of the many visible surfaces – the individual points in Figure 9a – is sufficient for making the correction, the correcting transformation still provides for the representation of the colors in the whole set in the low-dimensional representation.)

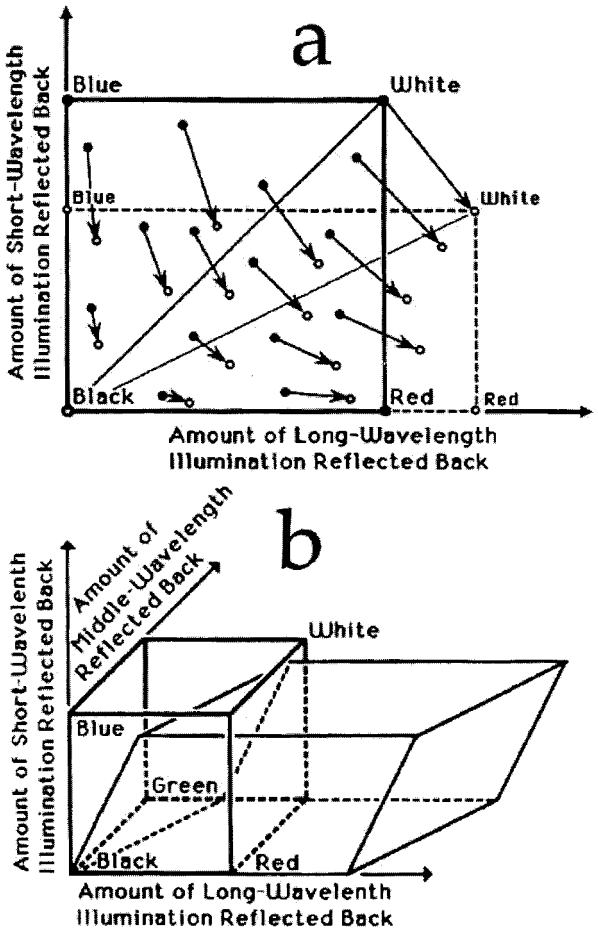



Figure 9. Schematic illustration of the effects of a terrestrial transformation on the amounts of light of different wavelengths scattered back to an eye: (a) for just two dimensions and a diagonal transformation, and (b) for a nondiagonal tranformation in three dimensions. (See text for explanation.)

The general case of a linear transformation that is both three-dimensional and nondiagonal is more difficult to illustrate for the whole set of points representing individual surfaces, but a rough idea of such a transformation may be gained from the more schematic depiction of the more general linear transformation between a cube and a parallelepiped in Figure 9b.

The dimensionality required is the same regardless of the particular transformation used to approximate the optimally color-constant transformation. This transformation could have the simplest (diagonal) form of the transformation described by Land and McCann (1971). It could have the more color-constant general linear form of the transformation proposed by Maloney and Wandell (1986; see also the revised approach described by Marimont & Wandell 1992). Or it could have some still more sophisticated form that takes account of surface orientation, shading, and shadows (see, e.g., Sinha & Adelson 1993); specular reflections (from glossy surfaces – see, e.g., Tominaga & Wandell 1989); or even, when the geometry permits the inference that the light falling on the object is identical to the light reaching the eye directly from the visible source, spectral properties of the illumination itself.

### 2.3. Formal characterization of the representation of invariant colors

As suggested by the preceding discussion of the spectral transformations of light by atmospheric filtering and surface reflection, and of the inverse transformations required to achieve invariance, these spectral transformations, like the rigid transformations of objects in space, constitute a mathematical group. Krantz (1975a; 1975b) has already presented an extensively developed group-theoretic formulation for the appearances of colored lights. From an evolutionary perspective, however, it was the invariant characteristics of light-reflecting objects – not the variable light or sources of light – that were of primary biological significance for the survival and reproduction of our ancestors in the pretechnological world. The linearity of the transformations of filtering and reflection ensures that the appropriate group for representing variations in the spectral composition of the light reaching our eyes from surrounding surfaces is a linear group, instead of the Euclidean group appropriate for rigid motions of objects in space. Of potential value, therefore, would be the further development of such a group-theoretic formulation of the representation of surface colors at a level of detail comparable to that provided in the group-theoretic representations of lights by Krantz (1975a; 1975b), of positions and motions by Carlton and Shepard (1990a; 1990b), of nonrigid deformations by Leyton (1992), or of musical intervals by Balzano (1980).

The formalization of the structures underlying psychological representation at a suitably abstract level can reveal deep analogies between disparate domains. In the domain


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

of color, just as in the domains of position, motion, deformation, and musical pitch, transformations have an abstract group-theoretic representation. Different domains require different groups, such as the Euclidean group for changes in position of an invariant shape, and the linear group for changes in the spectral composition of light reflected from an invariant surface. Nevertheless, they share some fundamental properties. In the representation of position or motion and in the representation of color, alike, the formal characterization reveals, for example, how prevailing structural constraints yield dimensional reduction of the representational space. Thus the symmetry group S(O) of a surface of revolution, such as a cylinder or a sphere, entails, through substitution of the appropriate quotient manifold (Eq. 2), a reduction from a six-dimensional to a five- or a three-dimensional space of distinguishable positions, respectively (Carlton & Shepard 1990b). Similarly, a restriction on the degrees of freedom of terrestrial filtering permits a reduction in the dimensionality of the representation for surface colors, from a space of six or more dimensions needed to capture the full reflectance characteristic of the surfaces of natural objects, to the three-dimensional space sufficient for the minimal invariant representation of their intrinsic colors.

### 2.4. Generality of the principles of color representation

Adaptation to the degrees of freedom of natural illumination does not of course ensure color constancy under conditions departing from those that have prevailed during terrestrial evolution. Modern technology has produced spectrally unnatural light sources under which human vision may not be color constant – as demonstrated in the vision laboratory, or under mercury vapor street lamps at night (where our companions may take on a ghastly aspect or we may fail to recognize our own car). Nor is an essentially three-dimensional variation of natural daylight the only factor that can determine the dimensionality of a species' color space. For nocturnal or deep-sea animals, the sensitivity afforded by achromatic rod vision may outweigh the benefits of cone-based color constancy. Even for many diurnal animals (including new world monkeys and human dichromats), the gain in color constancy attainable by the addition of a third class of wavelength-selective cones may be only marginal. Finally, runaway sexual selection may lead not only to the evolution of uniquely colored markings or plumage but also to the emergence of additional classes of retinal receptors and dimensions of color space tuned to the representation of such colors (Shepard 1992).

Still, the converging evolution of three-dimensional color representation in diverse visually dependent animals – evidently including most humans as well as the birds and the bees – may not be accidental. The speculation that I have favored is that this three-dimensionality may be an adaptation to a property that has long prevailed on our planet. We may need three dimensions of color not because the surfaces of objects vary in just three dimensions but because we must compensate for the three degrees of freedom of natural lighting in order to see a given surface as having the same intrinsic color regardless of that illumination.

The reduction specifically to three dimensions of color, though justified here in terms of the variations of natural illumination prevailing on earth, may hold more generally. On any planet capable of supporting visually advanced forms of life, illumination is likely to originate primarily from an essentially invariant stellar source. Moreover, the atmospheric and surface conditions necessary for such life are likely to provide only a limited spectral window of transmitted wavelengths of that light. Hence, the principal variations of the light reaching significant objects on or near the surface of such a planet are likely to be a variation in the overall quantity of that light and independent variations at the short-wavelength edge and the long-wavelength edge of the spectral window. Although additional, more subtle and spectrally selective variations may be to some extent present, these three variations – in the overall level of the transmitted light, and the extent of its reach into the short and the long wavelengths – seem likely to predominate in planetary environments generally and to exert the greatest influence through natural selection.

Whether or not my conjecture as to the nonarbitrary source and possible generality of the tendency toward three-dimensional color representation is ultimately supported, the universally linear way in which the spectral composition of light is transformed by scattering and filtering in the external world seems likely to have favored, wherever color vision has evolved, the internal implementation of compensating linear transformations on the proximal stimulus. Only in this way can significant external objects under any naturally varying illumination yield a color-constant internal representation, whatever its dimensionality may be.

## 3. Representation of an object's kind

The preceding examples concerned abilities to identify stimuli as distal objects that – despite wide variations in position and lighting – are nevertheless identical in intrinsic shape or color. My third and final example concerns an ability that does not require spatial or color vision and, hence, that is still more fundamental and ubiquitous. This is the ability to recognize that even when the distal objects themselves are not identical, they may nevertheless be objects of the same basic kind and, hence, likely to have the same significant consequences for the perceiver. For example, whether a newly encountered plant or animal is edible or poisonous depends on the hidden genetic makeup of the natural kind of that object.

Under the term *basic kind* I mean to subsume not only such *natural kinds* as animal, vegetable, and mineral species, but also such *basic level categories* (Rosch et al. 1976) as knife, bowl, or chair (for humans) or trail, burrow, or nest (for animals of some other species). Objects of the same basic kind are thus objects that provide the same functions or affordances (in the sense of Gibson 1979). A basic kind typically includes objects that, although more or less similar, may be readily discriminable from each other: an apple may be red or green; a trail may be level or steep; a chair may have a low or high back. Generalization from one object to another is not a failure of discrimination, therefore, but a cognitive act of deciding that two objects, even if readily distinguishable, may be similar enough to be of the same kind and, hence, to offer the same significant consequence or affordance.

This simple idea yields a quantitative explanation of a very general empirical regularity that is latent in generalization data of the sort that specifies, for all pairs of $n$ stimuli, the probability that a response learned to one of the



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

stimuli in the pair will be made upon presentation of the other stimulus of that pair. The latent regularity emerges when such data are submitted to multidimensional scaling, a method that finds the unique mapping of objects or stimuli into a space of minimum dimensionality such that the data have an invariant monotonic relation to corresponding distances in that space (see Kruskal 1964a; Shepard 1962a; 1962b; 1980). The resulting *generalization gradients,* which describe how the probability of a response learned to one stimulus falls off with the distance from it of each other stimulus in the obtained spatial representation, have uniformly approximated a function of simple exponential decay form. See Figures 10a and 10b, respectively, for generalization gradients that I obtained in this way (Shepard 1962b; 1965) for spectrally pure colors (hues) based on generalization data from pigeons (Guttman & Kalish 1956) and on a related type of similarity data from humans (Ekman 1954). The smooth curves are simple exponential decay functions fitted to the points by adjustment of a single parameter, the slope constant (which can equivalently be regarded as a scaling factor for the distances in each spatial representation). (For a number of gradients of generalization obtained in this way for other visual and auditory stimuli, see Shepard 1987b; and, for confirmation that the shape of the obtained functions is determined by the data and not by the multidimensional scaling method, see Shepard 1962b; 1965.)

Figure 11 displays the points corresponding to the spectral colors in the configuration (obtained from the human data) that yielded the approximately exponential function shown in Figure 10b. Four observations concerning this configuration are relevant here. First, as the close fit to the subsequently drawn circle indicates, the obtained configuration of points closely approximates Newton's color circle – the equatorial great circle of hues schematized in the earlier Figure 7a. Second, the implied psychophysical mapping from the rectilinear continuum of physical wavelengths to the circular continuum of perceived hues emerges as a consequence of my requirement that the law of generalization be not only invariant but monotonic – a requirement that was met, as can be seen in Figure 10b. (The pigeon data on which Figure 10a was based did not span a wide enough range of wavelengths to reveal this circularity.) Third, as I have already noted, the circularity of perceived hues is consistent with the opponent-processes representation of colors (Fig. 7b; see Shepard 1993; Shepard & Cooper 1992), which I conjecture to have arisen as an adaptation to the three degrees of freedom of natural illumination. Fourth, circular components, though historically ignored by most psychophysicists, arise in many representational spaces, including, in addition to those for color and for position and motion (considered here, and in Shepard 1978b), the chroma circle and the circle of fifths for musical pitch (Balzano 1980; Krumhansl & Kessler 1982; Shepard 1964b; 1965; 1982a; 1983), and the circadian, circa-lunar, and circannual components of time (e.g., see Enright 1972; Winfree 1980).

### 3.1. Formal characterization of generalization based on possible kinds

I originally derived the proposed universal law of generalization for the simplest case of an individual who, in the absence of advance knowledge about particular objects, encounters one such object and discovers it to have an important consequence. From such a learning event, the individual can conclude that all objects of this kind are consequential and that they therefore fall in some region of representational space that overlaps the point corresponding to the object already found to be consequential. Apart from its overlap with this one point, however, this *consequential region* remains of unknown location, size, and shape in representational space.

Although it is not essential to the basic theory, in the interest of keeping the initial formulation as sharp as possible, I propose for the present to proceed on the working hypothesis that the region in representational space corresponding to a basic kind is a *connected* region. Between the

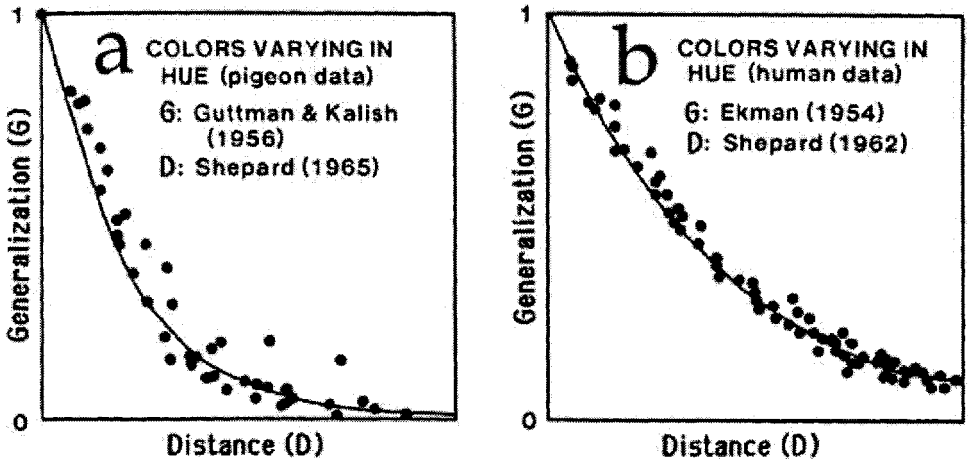



Figure 10. Generalization gradients for spectral hues obtained by applying multidimensional scaling to human and animal data: (a) based on the solution obtained by Shepard (1965) for the pigeon generalization data collected by Guttman and Kalish (1956), and (b) based on the solution obtained by Shepard (1962b) for the human similarity data collected by Ekman (1954). The distance, D, for each point is the Euclidean distance between the two colors in the multidimensional scaling solution based on generalization data, G; and the smooth curve in each plot is a one-parameter exponential decay function fitted by Shepard (1987b). From "Toward a Universal Law of Generalization for Psychological Science," by R. N. Shepard, 1987, *Science* 237:1318. Copyright 1987 by the American Association for the Advancement of Science. Adapted by permission.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

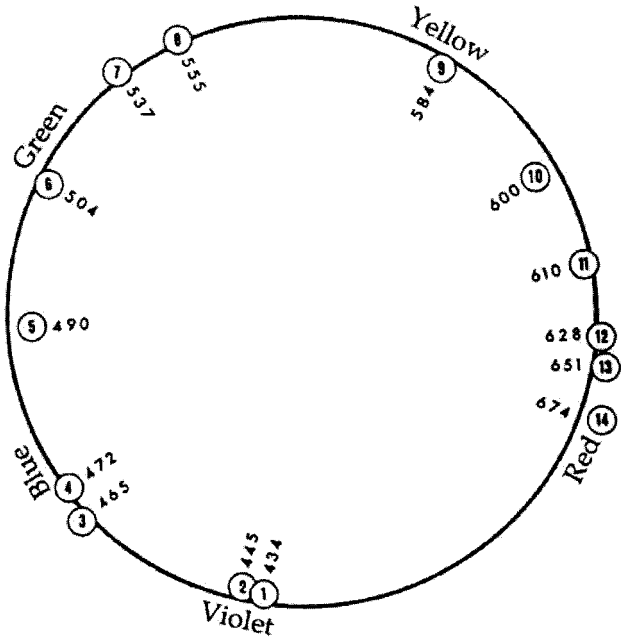



Figure 11. Multidimensional scaling configuration for Ekman's 14 spectral colors, obtained by Shepard (1962b) and corresponding to the plot shown in Figure 10b. The circle was subsequently drawn through the points to bring out the resemblance to Newton's color circle. The three-digit numbers indicate the wavelengths (in nanometers) of the corresponding stimuli. From "The Analysis of Proximities: Multidimensional Scaling With an Unknown Distance Function. II," by R. N. Shepard, 1962, *Psychometrika* 27:236. Copyright 1962 by the Psychometric Society. Adapted by permission.

points corresponding to any two objects of that kind, then, there is always a continuous path in the representational space that falls entirely within the consequential region for that kind. Thus, an apple could be continuously changed into any other apple, a chair could be continuously changed into any other chair, and a capital A could be continuously changed into any other capital A in such a way that at each step of the metamorphosis, the object would retain its recognizability as an apple, chair, or letter A, respectively.

The characterization of basic level categories in terms of a dichotomous criterion of connectedness, rather than in terms of some graded measure of correlation (of the sorts proposed by Rosch et al. 1976, and others) has two potential advantages: it can provide for the possibility of a sharp boundary between objects that though similar, belong to different natural kinds (only one of which, for example, manufactures a toxin); and it can provide for the possibility that objects of the same kind may nevertheless differ arbitrarily and widely in some of their features (an animal can vary enormously in size, shape, or coloration, for example, and still be a dog). Connectedness need not hold for nonbasic (e.g., superordinate or ad hoc) categories. There may be no continuous series between two such pieces of furniture as a sofa and a floor lamp for which every object along the way is also recognizable as a piece of furniture; and there may be no continuous series between two letters of the alphabet such as "B" and "C" for which every intermediate shape is also recognizable as a letter of the alphabet. Even for what I am calling basic kinds, my current working assumption of connectedness is only provisional.

I begin by considering an individual who has just found a particular, newly encountered object to have a significant consequence. This individual can only estimate the likelihood that a second, subsequently encountered object also has that consequence as the conditional probability that a random and (provisionally) connected region that happens to overlap the point corresponding to the first object, also overlaps the point corresponding to the second. The gradient of generalization then arises because a second object that is closer to the first in the representational space is more likely to fall within such a random region that happens to overlap the first.

To obtain an unbiased quantitative estimate of the probability that the new stimulus is consequential, the individual must use Bayesian inference. In effect, such an individual integrates over all candidate regions in representational space – with whatever prior probabilities, $p(s)$, are associated in that individual with the different possible sizes, $s$, for such regions. (In the absence of advance information, these prior probabilities are naturally assumed to be independent of the locations of the corresponding regions in the representational space.) For a test stimulus corresponding to a position $x$ in the representational space, the generalization $g(x)$ from a training stimulus (taken, without loss of generality, to be centered at the origin of an arbitrary coordinate system) to a new stimulus at location $x$ is then given by

$$g(x) = \int_0^\infty p(s)\frac{m(s,x)}{m(s)}\,ds, \tag{5}$$

where $m(s)$ denotes a (volumetric) measure of the region of size indexed by $s$, and $m(s,x)$ denotes a corresponding measure of the overlap between two regions of that size, one centered at $x$ and one centered at the training stimulus (i.e., at the origin).

The results of such integration turn out to depend remarkably little on the prior probabilities assigned (Shepard 1987b). For any choice of the probability density function $p(s)$ having finite expectation, integration yields a decreasing concave upward gradient of generalization. For any reasonable choice, integration yields, more specifically, an approximately exponential gradient. For the single most reasonable choice in the absence of any advance information about size – namely, the choice of the probability density function entailed by Bayesian inference from minimum knowledge or *maximum entropy* priors (see Jaynes 1978; Myung 1994) – integration yields exactly an exponential decay function (Shepard 1987b). Specifically, the maximum entropy assumption leads to a generalization function of the simple form

$$(\ ) = \left(--\right) \tag{6}$$

where the single parameter $k$ depends only on the expectation of $p(s)$.

Once again, invariance emerges only when formulated with respect to the appropriate, abstract representational space. To refer back to the domains of position, motion, and color, there is greater generalization between rectangles differing in orientation by 90° than between rectangles differing by somewhat less than 90°, and there is greater generalization between surfaces reflecting the shortest and longest visible wavelengths (violet and red) than between



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

either of these and a surface reflecting an intermediate wavelength (e.g., green). Clearly, generalization cannot be monotonic with distance in the usual physical space (of angle or wavelength). But generalization can become both invariant and monotonic with distance in the psychologically appropriate representational space, in which angles and wavelengths alike map into closed curves (Shepard 1962b; 1965; 1981a; Shepard & Farrell 1985).

Invariance in the law of generalization has thus been obtained by separating the *psychological* form of generalization in the appropriate psychological space from the *psychophysical* mapping from any specified physical parameter space to that psychological space. The psychophysical mapping, having been shaped by natural selection, would favor a mapping into a representational space in which regions that correspond to basic kinds, though differing widely in size and shape, have not, on average over evolutionary history, been systematically elongated or compressed in any particular directions or locations in the space. From what they learn about any newly encountered object, animals with a representational space for which biologically relevant kinds were consistently elongated or compressed in this way would tend to generalize too much or too little in certain directions of that space, relative to other species that had evolved an innate representational space that was appropriately regularized for the biologically relevant basic kinds in our world.

Ultimately, I expect the approach to generalization based on inference from maximum entropy priors, like the approaches I have already outlined for the representations of position, motion, shape, and color, to find a grounding in the theory of groups. This is because entropy, taken as a measure of the absence of knowledge (following Shannon 1948), can have a well-defined meaning only in relation to a space that (as I put it above) is properly "regularized" or that (in the words of Wiener 1948) has a "fundamental equipartition." To take the simplest example, if we have no knowledge about the location of a point in a one-dimensional space, we can only suppose that every location on the line is equally probable. (This is the "principle of indifference" so successfully employed in physics by Maxwell and Boltzmann – see Jaynes 1978.) Accordingly, the distribution that maximizes entropy in this case is, in fact, the uniform distribution. But if we were to transform this space by a nonlinear transformation (such as $x^* = x^2$ or $x^* = \log x$), what had been a uniform and maximum entropy distribution in the original space would no longer be so in the transformed space, and vice versa. Without going further into this deep and subtle matter here, I simply note that, in the opinion of one of the leading proponents of the maximum entropy approach in physics, "This problem is not completely solved today, although I believe we have made a good start on it in the principle of transformation groups" (Jaynes 1978).

### *3.2. Extensions of the generalization theory*

**3.2.1. Determinants of the metric of representational space.** A distinction that has been found basic to the understanding of similarity assessments and to discrimination and classification performances is the now widely recognized distinction between psychologically integral and separable relations among stimulus dimensions (e.g., see Garner 1974; Lockhead 1966; Shepard 1964a; 1991). This distinction has also been found to have a natural basis in the idea of consequential regions (Shepard 1987b; 1991; Shepard & Tenenbaum 1991). To the extent that the extensions of such regions along two or more dimensions have been positively correlated over evolutionary history, the integration over all possible regions, with their associated maximum-entropy weights, yields surfaces of equal generalization that approximate ellipsoids, implying the $L_2$ norm and associated Euclidean metric for that multi-dimensional representational space. To the extent that the extensions of such regions along the different dimensions have been uncorrelated, the integration over possible regions yields surfaces of equal generalization that approximate cross polytopes (a diamond-shaped rhomb in two dimensions, a triangular-faced octahedron in three), implying the $L_1$ norm and associated "city-block" metric for that subspace. In both of these multidimensional cases, integration still yields the exponential type of decay of generalization with distance in representational space originally derived for the one-dimensional case (for which the Euclidean and city-block metrics are equivalent). (The most appropriate group-theoretic representation is expected to be different, however, for conjunctions of integral and of separable dimensions.)

**3.2.2. Generalization over discrete features.** Although the derivation of the exponential gradient of generalization has been outlined here for the case of a continuous representational space, the theory is not restricted to the continuous case. When the objects possess only discrete (or even binary-valued) features, the analogs of the consequential regions in the continuous case become consequential subsets, and the analog of the volumetric size, $m(s)$, of a region becomes the (finite) number of objects in such a subset. Nevertheless, summation (the discrete analog of the integration used in the continuous case) still yields an exponential type of fall-off of generalization with distance, where distance is now defined in terms of the sum of the weights of the features that differ between the two objects or, if the features are all equal in weight, simply in terms of the number of differing features (Russell 1988; see also Gluck 1991; Shepard 1989).

**3.2.3. Classification learning.** Over a sequence of learning trials in which different objects are found to have or not to have a particular consequence, Bayesian revision of the prior probabilities associated with the various candidate regions yields a convergence to the true consequential region (Shepard & Kannappan 1991; Shepard & Tenenbaum 1991). Moreover, it does so in a way that agrees with results for human categorization (e.g., Nosofsky 1987; 1992; Shepard & Chang 1963; Shepard et al. 1961): The learning proceeds more rapidly when the consequential set of objects forms a region in the representational space that is connected rather than disconnected (Shepard & Kannappan 1991). The learning also proceeds more rapidly when the consequential set is compact in terms of the Euclidean metric if the dimensions are integral, but more rapidly when the consequential set is based on shared features (or conjunctions of features) if the dimensions are separable (Shepard & Tenenbaum 1991). (For related simulations, see Nosofsky et al. 1992; 1994; and for a similar Bayesian approach in which, however, the underlying hypotheses are taken to be Gaussian distributions rather than the sharply bounded regions posited here, see Anderson 1991.)

**3.2.4. A law of discriminative reaction time.** As I noted in the discussion of critical times in imagined and apparent


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

motion, natural selection has favored the ability to make decisions not only accurately but swiftly. But, whereas the time required to determine that two things are identical despite their apparent difference linearly increases with their transformational separation in the space of possible positions (as in mental rotation), the time required to determine that two things are different despite their apparent similarity nonlinearly decreases with their separation in the space of possible objects.

Specifically, latency of a discriminative response, like probability of generalization, falls off according to a decreasing, concave-upward function of distance between stimuli in representational space. But, whereas generalization probability, which cannot exceed one, approximates an exponential decay function of distance, discrimination latency, which is unbounded, is expected (under idealized conditions) to grow without limit as the difference between the stimuli approaches zero. In practice, such a function cannot be precisely determined for very small differences; experimental subjects would eventually either simply make a random guess or leave the experiment to terminate a potentially interminable trial. Nevertheless, functions that have been obtained do often approximate a reciprocal or hyperbolic form (e.g., Curtis et al. 1973; Shepard 1981a; 1989; see also Shepard et al. 1975). Such a form can be theoretically derived within the framework of the generalization theory. Suppose, for example, (1) that the internal representations corresponding to candidate regions overlapping either stimulus become activated, each with probability per unit time proportional to that region's associated prior probability of being consequential, and (2) that the first such representation to be activated – which overlaps one but not the other of the two stimuli – precipitates the discriminative response. Integration over all possibilities then yields, for the expected latency of discrimination, a reciprocal type of dependence on distance in representational space (see Shepard 1987b).

**3.2.5. The generality of generalization.** Presumably, things having the potential for particular, associated consequences belong to distinct kinds (including physical elements, chemical compounds, and biological species) and do so not just in the human or even the terrestrial environment but throughout the universe. If so, the exponential law of generalization, the reciprocal law of discriminative reaction time, and the Euclidean and city-block metrics of representational space may have arisen not just for the humans or animals we have studied on earth. Such laws and such metrics may have arisen wherever sufficiently advanced forms of life may have evolved. (This remains true even if biological species are themselves in part the product of mind – as suggested by the genetic algorithm simulations of Todd & Miller 1991.)

## 4. Conclusion

Perhaps psychological science need not limit itself to the description of empirical regularities observed in the behaviors of the particular, more or less accidental collection of humans or other animals currently accessible to our direct study. Possibly we can aspire to a science of mind that, by virtue of the evolutionary internalization of universal regularities in the world, partakes of some of the mathematical elegance and generality of theories of that world. The principles that have been most deeply internalized may reflect quite abstract features of the world, based as much (or possibly more) in geometry, probability, and group theory, as in specific, physical facts about concrete, material objects. By focusing on just three perceptual-cognitive examples – concerning the representation of the colors of objects, the kinds of objects, and the positions, motions, and shapes of objects – I have tried to indicate how psychological principles of invariant color, optimum generalization, and simplest motion may achieve universality, invariance, and mathematical elegance when formulated in terms of points, connected subsets of points, and geodesic paths in the appropriate abstract representational spaces.

ACKNOWLEDGMENTS

This article was originally published as the lead article in the first issue of the *Psychonomic Bulletin & Review* 1994, 1:2–28. The research and the drafting of this article were supported by the National Science Foundation (Grant No. DBS-9021648). Final revisions were supported, as well, by the Santa Fe Institute, and benefitted significantly from interdisciplinary discussions with many scientists there. Other colleagues who have especially contributed to or influenced the particular work reviewed here include Eloise Carlton, Lynn Cooper, Leda Cosmides, Joyce Farrell, Jennifer Freyd, Sherryl Judd, Laurence Maloney, Michael McBeath, Jacqueline Metzler, Geoffrey Miller, In Jae Myung, Robert Nosofsky, Margaret Shiffrar, Joshua Tenenbaum, Brian Wandell, and Susan Zare. Laurence Maloney and Vijoy Abraham greatly facilitated the conversion of the earlier version of the text and the figures to the present, edited electronic form, and Phineas de Thornley Head, of *Behavioral and Brain Sciences,* kindly entered a number of my final corrections to the electronic version.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 602–607
*Printed in the United States of America*

# The exploitation of regularities in the environment by the brain

**Horace Barlow**
*Physiological Laboratory, Cambridge CB2 3EG, England*
**hbb10@cam.ac.uk** **http://www.physiol.cam.ac.uk/staff/barlow**

**Abstract:** Statistical regularities of the environment are important for learning, memory, intelligence, inductive inference, and in fact, for any area of cognitive science where an information-processing brain promotes survival by exploiting them. This has been recognised by many of those interested in cognitive function, starting with Helmholtz, Mach, and Pearson, and continuing through Craik, Tolman, Attneave, and Brunswik. In the current era, many of us have begun to show how neural mechanisms exploit the regular statistical properties of natural images. Shepard proposed that the apparent trajectory of an object when seen successively at two positions results from internalising the rules of kinematic geometry, and although kinematic geometry is not statistical in nature, this is clearly a related idea. Here it is argued that Shepard's term, "internalisation," is insufficient because it is also necessary to derive an advantage from the process. Having mechanisms selectively sensitive to the spatio-temporal patterns of excitation commonly experienced when viewing moving objects would facilitate the detection, interpolation, and extrapolation of such motions, and might explain the twisting motions that are experienced. Although Shepard's explanation in terms of Chasles' rule seems doubtful, his theory and experiments illustrate that local twisting motions are needed for the analysis of moving objects and provoke thoughts about how they might be detected.

**Keywords:** Chasles' rule; evolution; geometry; perception; redundancy; statistics; twisting

## 1. Introduction

Statistical regularities abound in the world around us, and many of them are actually, or potentially, important for our survival. Furthermore, many of them are obviously exploited by our bodies, and the anatomy of the eye provides especially beautiful examples of evolutionary adaptations to different environments (Walls 1942). Shepard's papers (1984; 1994) were a major inspiration for the program at ZiF (Zentrum für interdisciplinäre Forschung) in Bielefeld that was the major source for this collection of articles. He evidently believed that the apparent motion trajectory of an object shown first at one position, then at another, resulted from the evolutionary adaptation of psychological mechanisms to the kinematic geometry of moving objects. Now, kinematic geometry is concerned with nonprobabilistic geometric relations, not statistics, but the way Shepard thought that perception was adapted to its rules parallels the way others have thought of perception adapting to statistical regularities. This article starts by giving a brief history of the development of these ideas, and then compares their predictions with Shepard's in the particular conditions he explored.

First, we should perhaps note that all learning could be regarded as the internalisation of environmental regularities, for it is driven by the statistically regular occurrence of reinforcement following particular sensory stimuli or self-initiated actions. Furthermore, it is well-recognised that statistical associations between sensory stimuli, as well as associations between the stimuli and reinforcement, influence learning (Mackintosh 1983; Rescorla & Wagner 1972). This makes it difficult to say where the use of statistical regularities in perception stops and their use in learning begins, but let us start with a brief historical review of the claims that have been made about their use in perception.

## 2. Helmholtz

Helmholtz flourished before Darwinian ideas about genetic adaptations to the environment were widely acknowledged, but he argued unremittingly that perception results from the interaction of *apperception* – the immediate impact of sensory messages – with *remembered ideas* resulting from past experience. He wrote of perceptions that " . . . by their peculiar nature they may be classed as conclusions, inductive conclusions unconsciously formed" (Helmholtz 1925). Thus, he held the view that experience of the environment was internalised or remembered and provided the basis, together with current sensory messages, for the statistical conclusions – mostly valid – that constitute our perception of the world.

## 3. Mach and Pearson

Ernst Mach (1886/1922) and Karl Pearson (1892) also appreciated the importance of environmental statistics, but

HORACE BARLOW, retired Royal Society Research Professor of Physiology at the University of Cambridge, has worked in Vision and Perception for 50 years. He contributed to the experimental discovery of bug-detectors and lateral inhibition in the frog retina, motion detection in the rabbit retina, and selectivity for disparity in the cat cortex. As a theoretician, he has argued that exploiting the redundancy of natural stimuli is important for effective perception and cognition, and that measuring absolute statistical efficiences is useful for defining the work involved.


© 2001 Cambridge University Press
0140-525X/01 $12.50
https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

they viewed the matter somewhat differently. They argued that scientific concepts and laws simplify our complex experience of the world, and that they are important because they bring "economy of thought" to our mental processes. Although this idea has great appeal at an intuitive level, it can only be made convincing if economy can be measured, and that had to wait for the quantitative definition of information and redundancy (Shannon & Weaver 1949).

## 4. Craik

In a short book, Craik (1943) developed the idea that the main function of the higher cognitive centres is to build symbolic working models. Such models must be based on the associative structure of objects and events in the environment, and are therefore expressions of environmental regularities. It is a more general form of Tolman's (1948) idea of cognitive maps.

## 5. Brunswik

Egon Brunswik (1956) seems to have been the first to suggest that the Gestalt laws governing grouping and segregation of figure from ground were more than empirical facts about perception: they were rules for using statistical facts about images to draw valid inferences from the scene immediately before the eyes. His work is not often quoted so it is worth describing in greater detail. He pointed out that objects have uniform properties compared with randomly selected regions of an image, and hence, if two patches have similar local characteristics, they are likely to be derived from the same object in the external world; this is the reason why it is appropriate to group them together. By analysing stills from the Alec Guiness movie "Kind Hearts and Coronets," Brunswik and Kamiya (1953) were able to show that there was a tendency for the proximity of two parallel lines to indicate the presence of a manipulable object in the scene, though this was a disappointingly weak effect.

The methods available then were feeble compared with those available now, and recent work on the statistics of natural images (Ruderman 1997) has shown that correlations of straightforward luminance values are indeed much stronger between points that lie within the same object than they are between points lying in different objects. Within the brain, images are not represented just by luminance values but by neurons selective for features such as orientation, texture, colour, disparity, and direction of motion. It will be interesting to see if the difference between inter- and intra-object correlations are even greater for these features than for luminance; if this is so, it would go a long way toward showing that the Gestalt laws of proximity, good continuation, common fate, and so on, are rules for making valid statistical inferences from environmental regularities. Elder and Goldberg (1998) have recently studied the validity of various properties of edges for bringing about correct object segregation.

It is worth noticing that Brunswik's idea makes a lot of sense of the anatomical arrangement of primary visual cortex (V1) and the surrounding extra-striate areas (Barlow 1981). V1 has neurons selectively sensitive to those local characteristics of the image that cause grouping, namely orientation, colour, texture, disparity, and direction of motion, most of which had already been identified by the Gestaltists. V1 neurons then project topographically to surrounding extra-striate areas creating new maps (see Lennie 1998), but there is also a non-topographic component in the projection. Neurons in V1 and V2 that are selectively sensitive to a particular feature (e.g., movement in a certain direction) converge on to single neurons in these extrastriate areas, thus collecting together information about this feature from relatively large regions of the visual field. Assembling the information in this way is the crucial step that enables such a feature to be detected at a low signal-to-noise ratio (Barlow & Tripathy 1997) even when it is spread over a fairly large patch of the image. Perhaps we are beginning to understand the physiology as well as the statistical logic of these first stages of object recognition.

## 6. Attneave

Attneave (1954) imported into psychology mathematical concepts that had been developed by Shannon and Weaver (1949) to quantify the transmission of information down communication channels. The most important of these from the present point of view are *information, channel capacity,* and *redundancy.* A communication channel can only transmit information at rates up to a finite limit called its capacity, but the messages actually transmitted often contain less than this amount of information; the difference is the redundancy of the messages. The importance here is that any form of regularity in the messages is a form of redundancy, and since information and capacity are quantitatively defined, so is redundancy, and we have a measure for the quantity of environmental regularities.

Attneave pointed out that there is much redundancy in natural images and suggested that the subjective prominence of borders provides an example of a psychological mechanism that takes advantage of this fact: you can represent an object more economically by signalling transitions between object and non-object because these are the unexpected, and therefore information-bearing, parts of the image. He illustrated with his famous picture of a sleeping cat that the same rule applies to the orientation of boundaries, for the picture was produced simply by connecting the major transition points in the direction of the border that outlines it.

## 7. Barlow

I became interested in the importance of statistical structure in sensory messages as soon as I came across Shannon's definitions of information, capacity, and redundancy. It seemed to me (Barlow 1959; 1961) that redundancy must be important throughout our sensory and perceptual system, from the earliest coding of physical messages by sensory receptors, right through to the intelligent interpretation of the patterns of excitation that occurs at the highest cognitive levels (Barlow 1983).

There has been one major change in my viewpoint. Initially I thought that economy was the main benefit to be derived from exploiting redundancy, as it is for AT&T and British Telecom. But, as explained in greater detail below, the physiological and anatomical facts do not fit the idea that the brain uses compressed, economical representations, and one can see that these would be highly inconvenient for many of the tasks it performs, such as detecting



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

associations. Therefore, I now think the principle is redundancy *exploitation,* rather than *reduction,* since performance can be improved by taking account of sensory redundancy in other ways than by coding the information onto channels of reduced capacity.

My initial idea was similar to Attneave's, described above, but what excited me was the fact that one could point to physiological mechanisms, such as the accommodation of sensory discharges to constant stimuli, light and dark adaptation, and lateral inhibition, that actually put the principles to work. I first wrote about it in 1956 (though the article was not published until 1961), and made predictions about the coding of motion that have subsequently been confirmed. If I had been smart enough I would have predicted the orientational selectivity of cortical neurons that Hubel and Wiesel (1959) discovered, for it has been shown that this fits the bill for redundancy exploitation (Ohlshausen & Field 1996; van Hateren & van der Schaaf 1998). An attractive feature of the idea is that a code formed in response to redundancies in the input would constitute a distributed memory of these regularities – one that is used automatically and does not require a separate recall mechanism.

The original article (Barlow 1961) suggested sparse coding, i.e., that the economy is brought about by reducing the frequency of impulses in neurons carrying the representation rather than by reducing the number of neurons involved. Barlow (1972) is mainly concerned with experimental evidence showing that single neurons in sensory pathways are highly sensitive and selective in their response properties; hence perceptual discriminations can be based very directly upon their activity and may characteristically depend upon only a few of the most active neurons. The article also develops the idea of sparse coding, where the activity of a small number of neurons selected from a very large population forms a distributed representation of the sensory input (see also Field 1994). The elements of this type of distributed representation are called "cardinal cells" to indicate their partial resemblance to Sherrington's "pontifical neurons." They signal directly the occurrence of messages belonging to subsets of the possible sensory inputs that it would be useful for an animal to learn about. The elements of distributed representations are often assumed to represent random or arbitrary subsets of input states, whereas a cardinal cell representation has some of the merits of grandmother or mother cell representations (see Lettvin's note in Barlow 1995), as well as those of sparse distributed representations.

Barlow (1989) argued for the general importance of the associative structure of sensory messages and proposed factorial coding, in which representative elements are formed that are statistically independent of each other, as a means of storing knowledge of these environmental regularities. Barlow (1990) suggested that motion and other after-effects result from adaptive mechanisms that tend to make representational elements independent of each other, and Carandini et al. (1997) provided some experimental evidence for the predicted contingent adaptation in neurons of monkey V1. Barlow (1996) reviewed some of this work and attempted to bring it up to date in the general context of Bayesian inference and perception.

The idea of economy in representation that Mach and Pearson proposed and Attneave and I recast in terms of Shannon's redundancy provides a key to understanding much in sensation, perception, and cognition, but the problem whether channel capacity decreases at higher levels in the brain needs to be faced. Initially it would seem that, if redundancy is to be reduced, the transformations in sensory pathways would have to generate very compact sensory representations with a reduced number of channels, each active for a high proportion of the time. In fact, almost the opposite occurs: at higher levels in the brain there are vastly more channels, though it is true that each is active at a lower rate. The increase in cell numbers is enormous, with more than a thousand times as many neurons concerned with vision in the human cortex as there are ganglion cells in the two retinas. The average frequency of impulses certainly becomes lower in the cortex, so coding does become sparser, but even if the capacity is deemed to be limited by this reduced mean firing rate, the increased number of cells dominates: on any plausible assumptions the capacity of the cortical representation is vastly greater than that of the retinal ganglion cells, so redundancy appears to be increased, not reduced. But as mentioned earlier, economising the capacity of the central representation may not be the factor of importance in the brain.

Information theory has always assumed that cost is proportional to channel capacity, and the commercial value of redundancy reduction lies in the reduction of costs it can bring about. Linking cost with channel capacity was appropriate for man-made communication channels, but costs and benefits may be quite different in the brain, and A. R. Gardner-Medwin and I have been looking at the "cost" for efficient learning that results from the use of distributed representations (Gardner-Medwin & Barlow 2001). Efficient learning requires the ability to count the occurrences of different attributes of sensory stimuli with reasonable accuracy, but unavoidable errors occur in distributed representations, where different attributes activate the same neuron. This overlap causes an increase in the variance of frequency estimates, which means that learning for features or events represented in a distributed manner must require the collection of more evidence than is necessary for features or events represented directly, in a localist manner. To reduce this loss of efficiency it is necessary to use many more neurons in distributed representations – that is, to *increase* their redundancy. This reduces the extent to which representational elements are active in events other than the one that is being counted.

Distributed representations allow an enormous number of different input patterns to be distinguished by relatively few neurons, and no one doubts that they are used in the brain. But for learning associations it is not sufficient just to distinguish different input patterns – one must also estimate how often they occur – and to do this with reasonable efficiency requires highly redundant representations. Perhaps this is a reason why there are so many neurons in the cortex.

Note that redundancy is a measure of any kind of statistical regularity, and there is no necessary relationship between the redundancy that can be exploited in the input, and the redundancy added by using a number of elements that is unnecessarily large for representation alone. Sensory redundancy is important because knowledge of regularities in the environment is advantageous for many purposes, such as making predictions. Redundancy in the representation has a quite different role: it reduces the extent to which elements are necessarily active for more than one type of input event, which is what hinders accurate count-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ing. There is no guarantee that the redundancy in the input would achieve the latter purpose, though it might be possible to transform input redundancy directly into a form that would reduce overlaps in the way required. Perhaps the "repulsion" between frequently co-active elements postulated to account for pattern-contingent after-effects (Barlow 1990) represents such a mechanism.

## 8. Watanabe

Watanabe (1960) drew attention to the similarity between inductive inference and recoding to reduce redundancy, a theme also taken up by Barlow (1974). If a particular type of regularity is identified in a mass of data, then it is possible to represent those data more compactly by exploiting the regularity. Carrying this argument to its logical conclusion, inductive inference is a matter of using statistical regularities to produce a shorter, more compact, description of a range of data, an idea that is carried further in the minimum description length (MDL) approach to these problems.

## 9. Minimum description length

Solomonoff (1964a; 1964b) and Wallace and Boulton (1968) suggested that the computer code with the minimum length necessary to reproduce a sequence of data provides the shortest, and therefore least redundant, representation of that data. This is obviously related to Occam's Razor, to Mach and Pearson's ideas about economy of thought, and to later ones on economy of impulses. The idea has been related to Bayesian interpolation (MacKay 1992), to the problem of simplicity and likelihood in perceptual organisation (Chater 1996), and to the general problem of pattern theory (Mumford 1996). Like redundancy exploitation, it uses regularities in the data to provide a basis for induction and prediction.

## 10. Recent work in this area

These ideas have become very much alive in the last decade or so. Linsker (1986a; 1986b; 1986c) applied information theory to understand the properties of neurons in the visual pathway, and Field (1987) tried to relate the properties of cortical neurons to the amplitude distribution for spatial frequencies in natural images. The statistics of such images were investigated by Tolhurst et al. (1992), Ruderman (1994), Field (1994), Baddeley (1996), and Baddeley et al. (1997). Atick (1992) reviewed the redundancy reduction idea, and Atick and Redlich (1990; 1992) argued that visual neurons were adapted to deal with the statistics of natural images. The idea that the form of the receptive field of V1 neurons is specifically adapted to the regularities of natural images has been around for a long time (e.g., Barrow 1987; Webber 1991) but early attempts were not very successful in using this principle to derive receptive fields like the real ones from the statistical properties of natural images. More recent attempts (Bell & Sejnowski 1995; Hyvarinen & Oja 1996; Olshausen & Field 1996) have used nonlinear methods and somewhat different principles. For instance, van Hateren and van der Scharff (1998) ran a program on natural images that performed Independent Component Analysis. This determines what receptive fields would be expected if the goal was to produce a limited number of descriptors of image patches that would, when added in the right proportions, generate accurate replicas of the range of images it was trained on. They showed that the predicted receptive fields match those determined experimentally in some, though not all, of their properties. More remains to be done along these lines, but it seems probable that the receptive fields of V1 neurons are indeed adapted to the regularities of natural images.

The work reviewed above suggests important roles for neurophysiological mechanisms that exploit the redundancy of sensory messages resulting from statistical regularities of the environment. For example, it has given us an idea why sensory nerves accomodate, why lateral inhibition occurs, why neurons are selectively sensitive to movement, why cortical neurons have the receptive fields they do, why and how the Gestalt segregation of figure from ground occurs, and why the striate and extrastriate visual cortex are organised the way they are. We now need to look more closely at Shepard's ideas to see how they are related to the adaptation of pre-perceptual mechanisms, through evolution and experience, to handle statistical regularities in the input with improved effectiveness. The next section argues that such adaptations could make perception expert in handling the images of moving objects, and that Shepard's idea of the internalisation of kinematic geometry, which emphasises nonprobabilistic geometric rules instead of statistical regularities, is too vague to describe the process and does not explain its advantages.

## 11. The problems of internalisation and kinematic geometry

Shepard's choice of the word "internalisation" is curious. In reading his original article it was not quite clear whether he thought this was straightforward evolutionary adaptation or not, but in his later (1994) article he clarifies this point by mentioning genes in the opening sentence, though he still frequently refers to internalisation. Now to understand the process as an evolutionary adaptation it is not sufficient just to copy the regularity internally, which is what the term internalisation implies. In addition, the regularity must be turned to some advantage, for without this the mechanisms would have no survival value. This is obvious in an example Shepard uses himself – diurnal rhythms – for a diurnal animal exploits the rhythm to become active by day and sleep at night while a nocturnal animal does the reverse, but both can be described as internalisation of the rhythm. In that case, only the appropriate phase has to be found in order to gain advantages, but for evolutionary adaptation to other environmental regularities the mechanisms required to gain advantages are likely to be more complicated and much less obvious. Let us try to apply this to Shepard's experiments.

When an object is shown successively in two positions, subjects experience it moving along a path between these positions, and Shepard claims that his experiments show the path to be close to that dictated by Chasles' rule. There are three problems here. First, Chasles' rule provides a concise way of describing how a three-dimensional (3-D) object can move from the first to the second position, but it does not say that the object has to move along the path corresponding to simultaneous translation and rotation, as



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Todorovič (this issue) explains in greater detail. As will be described below, it would be advantageous to have mechanisms adapted to respond to the types of motion that actually occur in the features of moving objects, and such mechanisms would be predicted on the redundancy exploitation hypothesis; but it is hard to see any basis for expecting these adaptations to correspond to the internalisation of Chasles' rule, because this does not necessarily specify the motions that actually occur.

Second, it is not clear whether the subject's judgements of intermediate positions are accurate enough to distinguish Shepard's predictions from ones based on mechanisms that respond to the different types of motion that occur frequently in the images of moving objects, as the redundancy exploitation hypothesis predicts.

The third problem is even more basic: we do not understand the neural basis for subjective experiences of moving objects, so it is risky to try to relate the experience to mechanism. To make sense of Shepard's claim we would have to accept a framework of the following sort. Assume the views of 3-D objects are represented in our brains by symbols for the coordinate values of prominent features in those objects. The views at the two positions would create two such representations, and we need to assume that the brain can interpolate between the corresponding features in these two representations to create representations of intermediate values. Shepard's claim would be that these interpolated representations correspond to positions along the screw transformation path that connects the two seen positions. With such a model the predictions of the hypothesis can at least be clearly stated, but it is a most implausible model because it assumes that experience is based on a temporal sequence of static representations, whereas we know there are neurons that represent movement: Where do these fit in? And how could the interpolated representations of static positions be formed early in the movement, before the object has appeared at its second position?

This model ignores modern knowledge of the neurophysiology of sensory systems and is very unconvincing, but one must have *some* model before one can make predictions. To go to other extremes, it is clear that there are some ways of representing image information in the brain for which it would be meaningless to talk of applying the rules of kinematic geometry. What if the 3-D object is described in our brains as "like a carrot bent in the middle," or in terms of the muscle activations required to place our finger on its various features? How could one apply the rules to these representations? The way perceptions are represented in our brains is far from settled, but the rules of kinematic geometry could not be applied to some possibilities.

Now, consider what advantages could be derived from having mechanisms evolutionarily adapted for signalling the motions of moving objects, for we need to understand these advantages if we are to relate the mechanisms to evolutionary survival. The results of van Hateren and van der Scharff (1998), briefly described above, suggest that there are cortical neurons that respond selectively to the spatial patterns that occur commonly in static natural images. If there are also ones that respond selectively to the spatio-temporal patterns of moving features in the images of moving objects, this would bring definite advantages. First, these neurons would act as matched filters for these patterns, and would therefore be optimal for detecting them at low signal-to-noise ratios, or as early as possible in the course of the movement. Second, they have the potential for extrapolating these movements into the future, that is for signalling a spatio-temporal pattern in its early stages, before it has been completed. And third, they could also interpolate, that is, signal the whole motion when only its first and last parts were actually visible; this capacity is obviously important in considering the interpretation of Shepard's experiments. Analogous advantages might possibly be obtained by applying kinematic geometry, but Shepard does not suggest these potential benefits. Furthermore, as Todorovič (this issue) points out, moving objects do not necessarily follow the helical path described by Chasles' rule, so its predictions might be misleading; the ordinary redundancy exploitation hypothesis makes more sense.

## 12. Geometric rules, the rigidity assumption, and experience

Shepard bravely proposed that a geometric rule, not a statistical regularity, is internalised, and that this leads directly to predictable subjective experiences. Sinha and Poggio (1996) have recently described experiments that show how subjective experience is influenced by interactions between the mathematical rules of perspective transformations, recent experience of particular motions, and a tendency to interpret motions that occur together as motions of a rigid object. Though the details are quite different, the important factors in these experiments are sufficiently similar to those in Shepard's experiments for their results to be relevant here.

Sinha and Poggio's subjects look for a few minutes at the computer-simulated silhouette of a 3-D figure made of straight wire segments being rotated to and fro about a horizontal axis. This is normally perceived as a rigid 3-D body being rotated, and if tested with the same figure and the same motions within a few minutes, the impression of rigid rotation is retained. But if instead a silhouette that is identical at its mid-position is moved in a way corresponding to a wire-frame figure that has a different 3-D shape, then the subjects frequently perceive nonrigid motion: that is, the object appears to deform as it moves. If those same motions had been seen without the training experience, they would have been perceived as the rigid rotation of a different 3-D body. Therefore the subjects certainly remember or internalise something as a result of their initial experience, but it is not the laws of kinematic geometry: it is that particular 3-D shape of the wire-frame object seen in the initial adapting experience that would, when rotated, generate the set of images that was actually experienced.

I think Helmholtz would have been delighted by this experiment, not only because it illustrates so well the relation between his "apperception" and "remembered experience," but also because it brings out something that would have been new to him. It is the initial assumption of rigidity that makes it logically possible to infer a 3-D shape from the rotating image, and this gives new insight into the role of such a "default assumption" in perception. Furthermore, this type of assumption is presumably genetically determined and is the consequence of evolutionary selection, both of which would, I think, have been further new ideas for Helmholtz.

If I have understood Shepard correctly, he thinks that perceptions somehow embody as a whole the regularities constituting kinematic geometry. In contrast, this experi-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ment shows how the laws of perspective transformation are used, together with the rigidity assumption, to form one particular detail of the percept, namely its 3-D shape. This fits the ideas reviewed in previous sections of this article: regularities in the motions of the wire segments are detected and used to construct a rigid 3-D shape that is compatible with them and, when possible, with the rigidity assumption; it is this shape that we experience and which influences subsequent interpretations of motion. A hierarchy of operations occurs in the visual pathways, and at least for the early ones the evidence is now strong that they conform to the principle of exploiting redundancy. In this light, the idea that perception has internalised the rules of kinematic geometry seems vague and implausible. Furthermore, it is doubtful if Shepard's experimental tests of his specific idea distinguish it from the more general hypothesis about exploiting statistical regularities.

## 13. Conclusions

The principle that the redundancy in sensory messages resulting from regularities in the environment are exploited in sensory pathways illuminates a host of sensory phenomena, such as accomodation, light and dark adaptation, lateral inhibition, the form of feature detectors in the cortex, their relation to the Gestalt laws, the organisation of extrastriate areas, the functional role of figural and contingent after effects, and possibly the nature of intelligence itself. The principle of adaptation to regularities has a very respectable past, it is a fertile inspiration for current research, and looks set for a prosperous future. But Shepard has certainly drawn attention to an interesting phenomenon, and there may be an important lesson for neurophysiologists to learn from it.

Shepard claimed that when an object is presented first in one position, then another, " . . . one tends to experience that unique, minimum, twisting motion prescribed by kinematic geometry." There may be doubts about the role of kinematic geometry, but there is certainly a rotary component to the motion experienced, and once it has been pointed out it is clear that such rotations must play an important part in the interpretation of the images of moving objects. Furthermore these rotations need to be tightly localised to particular image features when using them in this way, and V1 is the high resolution area of the visual cortex. Hence, one must ask "Do neurons in V1 detect rotation directly?" A recent analysis of their responses at different delays after the presentation of an oriented stimulus found that, for some of them, the favoured orientation does in fact change with the delay (Ringach et al. 1997), so the optimum stimulus would be a twisting motion. Since Hubel and Wiesel we have known that V1 neurons signal the orientation of edges in the visual field: do they also signal *change* of orientation with time? This hint that some V1 neurons are tuned to twisting movements, not pure translations, urgently needs to be followed up.

ACKNOWLEDGMENT

The author was supported by Grant No.046736 from the Wellcome Trust during the preparation of this paper.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 608–617
*Printed in the United States of America*

# Regularities of the physical world and the absence of their internalization

**Heiko Hecht**
*Man-Vehicle Lab, Massachusetts Institute of Technology, Cambridge, MA 02139*
**hecht@mit.edu** **www.uni-bielefeld.de/zif/heiko.html**

**Abstract:** The notion of internalization put forth by Roger Shepard continues to be appealing and challenging. He suggests that we have internalized, during our evolutionary development, environmental regularities, or constraints. Internalization solves one of the hardest problems of perceptual psychology: the underspecification problem. That is the problem of how well-defined perceptual experience is generated from the often ambiguous and incomplete sensory stimulation. Yet, the notion of internalization creates new problems that may outweigh the solution of the underspecification problem. To support this claim, I first examine the concept of internalization, breaking it down into several distinct interpretations. These range from well-resolved dynamic regularities to ill-resolved statistical regularities. As a function of the interpretation the researcher selects, an empirical test of the internalization hypothesis may be straightforward or it may become virtually impossible. I then attempt to cover the range of interpretations by drawing on examples from different domains of visual event perception. Unfortunately, the experimental tests regarding most candidate regularities, such as gravitational acceleration, fail to support the concept of internalization. This suggests that narrow interpretations of the concept should be given up in favor of more abstract interpretations. However, the latter are not easily amenable to empirical testing. There is nonetheless a way to test these abstract interpretations by contrasting internalization with the opposite concept: externalization of body dynamics. I summarize evidence for such a projection of body constraints onto external objects. Based on the combined evidence of well-resolved and ill-resolved regularities, the value of the notion of internalization has to be reassessed.

**Keywords:** event perception; evolution; internalization

## Introduction

Shepard's (1994) claim that our minds reflect the very same principles that govern the universe is appealing indeed. According to this claim, the mind has internalized universal principles (regularities) that allow it to disambiguate situations that would otherwise be unsolvable. Provided the world is not changing, such universal principles are very efficient. For the visual system, this explains why we can make sense of stimuli that by themselves do not suffice to specify our perceptions. As I will show in this paper, as appealing as this claim is, it has two interpretations that need to be distinguished. Both become problematic when subjected to closer scrutiny.

There is a troubling duality to Shepard's internalization hypothesis. On the one hand, the convincing example of an inner circadian rhythm suggests he takes internalization to mean that a well-defined physical regularity is also independently present in the organism and allows behavior consistent with the regularity even if it is no longer there (as is day or night for people in a dark cave). This example is quite unique, and other examples, such as kinematic geometry, do not assume any exact mirroring of a physical law in the perceptual or behavioral outcome. To the contrary, kinematic geometry supposes a good deal of abstraction from movements that are found in the physical world. The two examples are symptomatic for two vastly different readings of the internalization hypothesis. The former I call the literal interpretation. The latter I call the abstraction interpretation.

For the literal interpretation, to determine whether or not some principle or regularity of the physical world has been internalized, three things have to be true: (1) First, there has to be a regularity in the world that can be assessed independently of our perceptions. (2) Second, our behavior and/or percept has to be compatible with this regularity as established by empirical observation. (3) Third, additional evidence is needed to show that the percept has come about by virtue of internalization and not by some other learning process. The first two steps are comparatively straightforward while the third is very tricky. Fortunately, it does not have to be resolved when approaching the problem from a falsificationist point of view. As long as one and

HEIKO HECHT is a post-doctoral associate working in the Man-Vehicle Laboratory at the Massachusetts Institute of Technology, Cambridge, MA. He has published on issues of motion perception and dynamic event perception. From 1996–1997 he was scientific coordinator of an international project on Evolutionarily Internalized Regularities at the Center for Interdisciplinary Research in Bielefeld, Germany.


© 2001 Cambridge University Press
0140-525X/01 $12.50
https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

two are the case, the (literal) internalization hypothesis survives the test.

A less literal reading of the internalization hypothesis causes much more trouble. It is also what Shepard most likely had in mind. To arrive at perceptual-cognitive universals (Shepard 1994), a mere copy of an often seen movement or event is clearly insufficient. Internalization as a process of abstraction is geared toward finding default solutions. Whenever the stimulus is ambiguous or ill-defined, as in apparent motion, an internalized default influences the percept. In this evolutionary process, geometry has been more deeply internalized than physics (Shepard 1994). Thus, an almost paradoxical relationship between the degree of internalization and its palpability is postulated. The deeper an invariant is internalized the more abstract it has to be. The process of internalization is then not just an inductive process, but also, by definition, one where the best examples are the least well-defined.

Unfortunately, this abstraction interpretation virtually annihilates the three requirements that hold for the literal interpretation: (1) The physical regularity no longer needs to be crisp and stateable as a physical law. (2) It becomes much harder to state what empirical behavior actually contradicts an abstract internalized rule. Together, this leads to potential research problems that are addressed below. (3) With the broader interpretation of the internalization hypothesis, the issue of internalization versus learning may reach beyond what can be tested empirically (see Schwartz, this issue).

An empirical assessment of Shepard's internalization hypothesis – and this is my quest – is thus inextricably tied to its interpretation. The literal and the abstract interpretation can be taken to represent the two ends of a continuum, within which Shepard is hard to place. Because of this difficulty, I resort to the strategy of evaluating a variety of internalization candidates that range from very narrow to very broad readings of internalization. For all of them I stick to the general premise that the internalized knowledge comes into play when the percept is ill-defined or when conflicting cues have to be resolved. Since a hypothesis should not be tested with the examples that were drawn up for its initial support, I pick some domains that have not been considered by Shepard but in my opinion constitute good cases for potential internalization. These are some literal laws of physics as well as some examples from the more abstract domain of intuitive physics. I hope thus to analyze Shepard's claims and to elucidate the concept of internalization.

## 1. Classifying candidates for internalized regularities that govern the physical world

A natural strategy for an empirical test of *the* internalization hypothesis would be to first examine different types of regularities at differing levels of abstraction, and then to test whether our percepts reflect these regularities. To evaluate a given candidate regularity, three questions should be answered. First, to what degree does it describe the physical world, that is, are there exceptions or is it universally true? Second, what is its level of complexity? A very complex natural law may hold without exception but it might be impenetrable to the visual system and appear inconsistent. Third, what is the degree of abstraction that is involved in a given internalization hypothesis? Possible candidates can be grouped as a function of how they score on these questions. I distinguish the following groups: potentially internal regularities that are close to the laws of physics, such as dynamic invariants; specific but highly abstracted rules, such as kinematic geometry; more general rules, such as the Gestalt principles; and unspecific and highly abstracted regularities, such as Bayesian probabilities.

### 1.1. Dynamic regularities

The strongest case for the internalization of physical regularities would be made if a simple invariant that holds in the physical world guides our perception. The fact that light usually comes from above seems to fit this category perfectly. There are many examples of unexpected and unnoticed artificial illumination from below, however, that can perceptually invert the scene. Valleys are turned into mountains and vice versa (Metzger 1975; Ramachandran 1988a). Light does not always come from above, though *generally* it does and this may have prompted the visual system to use that assumption when the stimulus is not very rich, as when looking at photographs or masks of human faces. This "illusion" has not been reported in more ecological settings, but that poses no threat to the internalization idea. At this literal level, only the failure to recur to plausible regularity assumptions would pose a threat to the hypothesis. I contend that such data is there to be used and can be gleaned from studies of intuitive physics (see the section on candidates for internalization). Unfortunately, the other good candidates, such as the constant gravitational acceleration of falling bodies, do not seem to support the notion of successful internalization.

### 1.2. Geometric regularities

The most detailed internalization hypothesis that Shepard (1984) has put forth is that of kinematic geometry. He proposed the internalization of geometric principles pertaining to group theory at a high level of abstraction. These principles, which prescribe, among others, circularly curved motion paths, are thought to act as a general default that influences perceiving, imagining, and dreaming. This abstraction variant of the internalization hypothesis remains very controversial (see Todorovič, this issue). The empirical evidence gathered by Shepard himself (e.g., Lakatos & Shepard 1997) causes confusion about what exactly is meant to be internalized. Three different views are possible and leave a number of back doors open to maintain abstract internalization: (1) The crisp law of geodesic movements could have been "imperfectly" internalized. (2) A general, imperfect law could have been perfectly internalized. (3) A fuzzy general law could have been imperfectly internalized. It is not hard to see that empirical data can be imperfect in multiple ways and still be compatible with Shepard's proposition. This issue will be taken up in the section on kinematic geometry.

Another example of how the visual system exploits knowledge about geometric regularities has been put forth by Bingham (1993). He found that observers use the shape of unfamiliar trees to judge their absolute size. To do so general relationships such as ratio of trunk to branch size, number of branches, and so on are exploited. Here shape can even override horizon-ratio information, which is normally very informative (Rogers 1995), at least as far as pictures are



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

concerned. Thus, there is evidence outside the realm of kinematic geometry demonstrating that observers make use of prior shape knowledge to judge absolute size. However, this potential role of geometric regularities in perception does not entail that the geometric knowledge is internalized (vs. learned), or universal.

#### 1.3. Gestalt principles

A next step of abstraction is reached when Gestalt principles are interpreted as internalized regularities. Gestalt psychology shares the conviction that the different Gestalt principles reflect very general regularities. The well-known Gestalt principles, such as grouping by similarity, by common fate, or, more recently, by uniform connectedness (Palmer & Rock 1994) will not be discussed here because traditionally they have not touched on the issue of internalization, probably because Gestalt psychologists were more concerned with a static description of phenomena and physiology than with evolutionary processes. Some Gestalt theorists (e.g., Metzger 1975) have even treated Gestalt principles as the very conditions that make perceptual psychology possible. Thus, Gestalt psychologists acknowledge the pervasiveness and *a priori* nature of a whole list of principles and they do not single out one, such as kinematic geometry. They also do not stress the processes of internalization but conceive of them in an almost Kantian fashion as preconditions of experience. The notion of an internalization could be taken as an evolutionary explanation of the origin of general principles, including Gestalt principles.

#### 1.4. Statistical regularities as a special case of maximal abstraction

The most abstract way to describe regularities of the physical world that are reflected in the visual system consists in pointing out mere statistical relationships. Typically, if an equal distribution assumption is made, we can predict which views of objects are likely and which ones are rare. For example, it is extremely unlikely that we see a pencil exactly head-on such that it produces the retinal image of a circular patch (provided by monocular viewing). Consequently, a circular retinal patch is normally not interpreted as a pencil but rather as a round object. The notion that the visual system "knows" generic views from accidental ones has been put forth in Bayesian approaches to perception (e.g., Albert & Hoffman 1995; Hoffman 1998), which postulate that the organism makes use of prior information about the world. For instance, Hoffman (1998) describes such knowledge as a list of rules that the visual system applies to the stimulus, such as "interpret[ing] a straight line in an image as straight line in 3-D."

This reconstructionist view gathers support from Shepard (1987b), who suggests how such prior knowledge could have developed by a process of internalization. His explanation draws on probabilistic aspects of nature and processes of stimulus generalization within the organism. The likelihood of responding to a new stimulus the same way as to a different previously learned stimulus (generalization) depends on the proximity between the two stimuli in psychological space. According to Shepard, this function is not equivalent with discriminability but reflects the anticipated consequences of the reaction toward the stimulus class. The function is exponential and supposedly reflects a universal law that is as ubiquitous for animate beings as Newton's law of gravitation is for inanimate objects. This is a good argument for why internalized laws are poorly resolved and may have to be imprecise. Unfortunately, it makes it very hard to interpret empirical data that do not quite fit the supposed regularity. On the one hand, such data could be taken to mean that a well-resolved regularity has been internalized poorly. On the other hand, it could mean that the regularity has been abstracted and then internalized perfectly. Without a set of independently derived abstraction rules, we cannot favor one interpretation. And since such rules have not been formulated, the internalizationist's foregone conclusion is that internalization has been demonstrated. This seems to prompt Proffitt and Kaiser (1998) to conclude that the visual system has not internalized (well resolved) dynamic constraints but rather (coarser) geometric concepts.

It is easy to arrive at this conclusion under the premise that internalization has to be perfect, but this is most likely not the case. If this be demanded, existing empirical evidence suffices to falsify claims of internalization of both dynamic and geometric concepts. To support this point, I will summarize representative empirical evidence showing that our percepts are often only approximated by such concepts. Neither dynamic invariants (gravity, horizontality) nor optical invariants (tau) nor geometric rules (geodesic paths) predict our perceptions with satisfactory accuracy. Shepard has tried to turn this vice into a virtue by introducing a process of abstraction into the concept.

## 2. Three example cases for internalization

If universal but specific regularities can be found which appear to guide our perception whenever underspecified, a case could be made for internal knowledge and maybe even for a process of internalization. Once this is done, more abstract, generalized versions of physical regularities can be considered. Thus, I first examine gravity and horizontality as potential dynamic regularities. Based on the negative results, the would-be universality of apparent motion trajectories will then be reconsidered, reevaluating the example of kinematic geometry.

#### 2.1. Gravity as a cue to absolute size and distance

The force of gravity is not only ubiquitous but also accelerates all terrestrial objects at a constant rate. Gravity is thus a prime candidate for a specific constraint that the visual system might have internalized to disambiguate perception. The internal knowledge in this case would be indirect. If observers judge the absolute size of an object more accurately when they see it fall, they may use implicit knowledge about gravitational acceleration to perform this task. Saxberg (1987) and Watson et al. (1992) suggested that observers do in fact estimate the absolute size and/or distance of objects by relying on the monocular cue of gravitational acceleration as is present in projectiles in flight, pendulum motion, fluid wave motion, and others. Saxberg, for example, showed that one could estimate the absolute distance to an object from four retinal image variables: the vertical and horizontal components of the object's retinal velocity and the vertical and horizontal components of its retinal acceleration. The estimation of absolute distance is even sim-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

pler when the object's motion is vertical; the vertical retinal acceleration in that case is proportional to the gravitational constant (Watson et al. 1992). Thus, one can in principle estimate the absolute distance to a freely-falling object. This estimate becomes more complex when friction plays a role as is the case for light and fast objects. However, for most inanimate objects within our space of action, air resistance has comparatively small effects.

Empirical results, however, showed that observers do not behave as if they make use of some knowledge about gravity (Hecht et al. 1996). Computer-simulated events of free falling objects revealed that observers were not very good at scaling absolute size and/or distance. Balls of different diameters and at different distances from the observer were simulated to rise, climb to their apex position, and then fall back down. Two categories of events were used, accelerating balls and constant-velocity balls. The latter had the same event-durations and average velocities as the former; however, only accelerating stimuli could be used to scale the distance of the event. Figure 1 depicts the different position/time diagrams for a subset of the stimuli whose distance observers had to judge. Observers did not perform better on the accelerating trials than on constant velocity trials, but both were considerably better than static versions of the stimuli. It can thus be ruled out that observers used projected size as a cue, which is always correlated negatively with simulated distance and positively correlated with simulated size.

Thus, observers do not utilize specific knowledge about gravitational acceleration to a sufficient degree. It remains possible that some abstraction of this regularity has nonetheless been internalized. Average image velocity is necessarily and negatively correlated with simulated distance whenever the apex point is shown. The fact that size and distance judgments in the constant velocity condition were significantly better than chance shows that observers can, in fact, make use of the average velocity cue. They behave as if they were abiding by a simple heuristic such as "objects that produce fast retinal motion are relatively close to me."

The hypothesis that gravitational acceleration has nonetheless been internalized could be salvaged by assuming that, for some reasons, fast-moving objects for whom air resistance is no longer negligible, have determined the internalization process. In this case, air-resistance is sometimes considerable and the effects of gravity vary depending on density and size of the falling objects. Drag, for instance, increases geometrically with object velocity. For a baseball moving at 80 miles/hr, the drag is about 70% of the ball's weight (Brancazio 1985). Thus, the visual system, instead of having to adjust for drag, might have adopted a cruder mechanism reflecting the fact that moving/falling objects give rise to higher retinal image velocities at closer distances. This relationship usually holds no matter how the object moves and whether the object is accelerating or moving at constant velocity. Presumably, observers are sensitive to this fundamental relationship and the visual system could use this abstract information to disambiguate percepts of distance.

In sum, we have to reject the falsifiable hypothesis that observers have internalized detailed knowledge about the rate of gravitational acceleration. The less specific case is still possible, but it may also be immune to criticism (see Fig. 4).

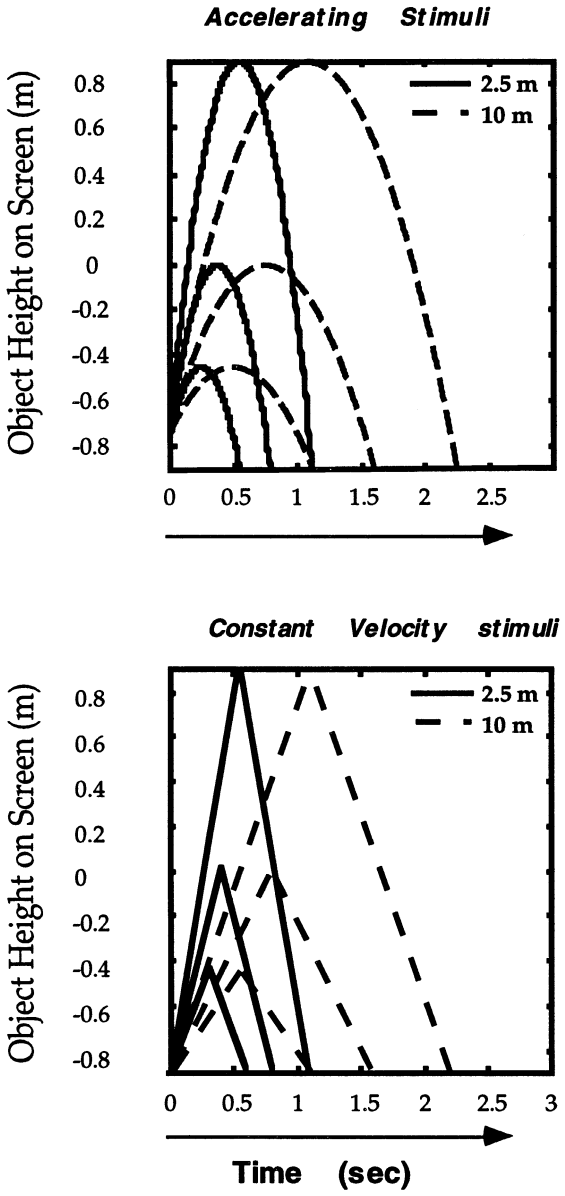



Figure 1. Projected ball trajectories for accelerating and constant velocity trials. Vertical position on the display screen is plotted as a function of time. The left panel shows the trajectories for the accelerating condition for simulated distances of 2.5 and 10 m. There were three different apices in this experiment, one at the top of the screen, one three-quarters of the way to the top, and one half of the way. The right panel shows the trajectories for the constant velocity condition for simulated distances of 2.5 and 10 m (adapted from Hecht et al. 1996, p. 1070).

### 2.2. The law of horizontality and the water-level task

When shown a tilted container people often fail to appreciate that the surface of the contained liquid should remain horizontal with respect to the ground. A typical example of the paper-and-pencil version of this Piagetian task is depicted in Figure 2. Subjects are asked to draw in the surface of the water such that it touches the dot on the right side of the container. About 40% of the adult population draw water-levels that deviate by more then 5% ($\alpha$) from horizontal (for an overview see Liben 1991). The failure to solve this water-level task correctly is quite robust



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

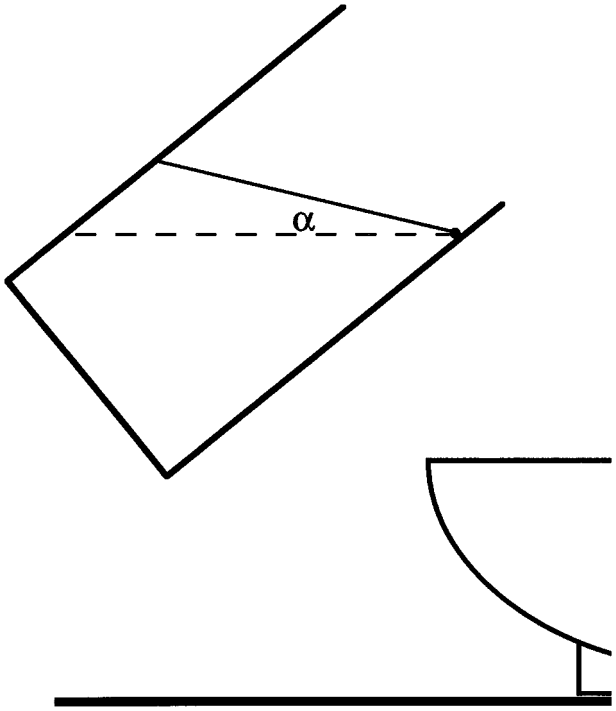



Figure 2. The water-level task. Observers have to draw the surface of the water. They are told that the beaker is at rest and filled by as much water as needed to make the surface touch the dot on the right of the container. The dotted line indicates the correct solution, which was only produced by one-half of all subjects. The solid line depicts a typical answer given by the other subjects.

across presentation contexts and does not appear to be an artifact of the technique that is chosen to communicate the task.

Howard (1978) presented apparent motion sequences of photographs depicting horizontal and oblique water-levels and asked subjects to report whether the sequence represented a natural or an unnatural event. Using an animated version of the task did not improve performance (Howard 1978; McAfee & Proffitt 1991). The would-be internalization of horizontality did not come to the fore when pouring events with impossible water-levels were shown by virtue of tilting the camera when the scenes were videotaped. Thus, a variety of methods used to assess the explicit and implicit knowledge of the horizontality invariant produced the same results: the regularity that liquid surfaces at rest remain invariably horizontal with respect to the ground cannot be taken to be internalized by our visual system.

If the horizontality of liquids has been internalized in a more ephemeral manner, visual experience may be required before the internalized regularity manifests itself in behavioral data. Thus, one might argue that with sufficient experience the "illusion" should disappear. However, the opposite is the case. Experienced waitresses and bartenders reveal stronger biases than the average population; they accept water levels as natural that deviate even more from horizontality (Hecht & Proffitt 1995; but see Vasta et al. 1997).

In conclusion, two examples of physical regularities have failed to influence perceptual judgments. The empirical data have thereby failed to fulfill a precondition for the possible internalization of theses regularities. Percepts were vastly different from the defaults that hold in the environment with few exceptions. Knowledge about gravity, per se, is not used to scale absolute size and distance of objects. However, the general negative relationship between retinal velocity and distance of a moving object could still be said to be internalized. A much stronger case against the internalization hypothesis is represented by the water-level literature. The regularity that liquid surfaces remain invariably horizontal when at rest is as consistent as the diurnal cycles, it has no exceptions. Nonetheless, a substantial proportion of observers misjudge water-levels, indicating that they have by no means internalized this particular regularity.

Thus, the failure to exploit these rather concrete regularities contradicts the literal reading of Shepard's hypothesis. However, neither free-fall nor the water-level task have a geometric solution that differs from the laws of classical mechanics. Thus, if one excludes these examples from the domain of internalization theory, the latter may not be threatened. Such an exclusion might be put forth on grounds of insufficient underspecification of the percept in the case of gravity to scale objects and on the importance of cognitive factors in the water-level case. This would rescue the literal interpretation but seriously reduce the scope of the hypothesis.

### 2.3. Kinematic geometry?

Let us now look at the abstraction reading of internalization. Given the negative results that were obtained when testing the comparatively simple regularities of gravity and horizontality, we search for evidence of internalization in situations where the percept is severely under-specified but not arbitrary. Such cases are very hard to find. Dreaming and imagining – although suggested by Shepard (1984) – appear to lack specification altogether and are, in addition, very hard to measure. Apparent motion seems to be the only appropriate case that could provide evidence for internalization of abstracted regularities.

For moving extended objects, Shepard (1984; 1994) claims that the perceived trajectory of an object's motion is determined by the geometrically simple geodesics. The model is based on the idea that a group of single rotations (SO[3]) can define the space of all possible three-dimensional (3-D) orientation differences (Carlton & Shepard 1990a; 1990b; Foster 1975b). Within this space, Chasles' theorem describes the simplest single rotation as follows: for any object displacement and orientation change, there exists one axis in space about which the object can be rotated, such that its initial position will be mapped into its final position. This helical motion in 3-D reduces to a single rotation (without concomitant translation) in the 2-D case.

However, the empirical evidence, including some of Shepard's own studies (McBeath & Shepard 1989), does not always support the geodesic model. McBeath and Shepard's data fell somewhere between the straight line path suggested by principles of energy minimization and the postulated geodesic path. Empirical apparent motion trajectories in 3-D especially often deviate considerably from the geodesic solution. Depending on the circumstances, perceived paths can be much closer to a straight line than to a geodesic curve even when the 3-D orientation of the motion plane necessary to specify the geodesic is properly judged (Hecht & Proffitt 1991). These results hold, of course, only when inter-stimulus intervals are sufficiently long so that geodesics


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

could in principle be observed. A general model not based on kinematic geometry that could explain many of these results has been proposed by Caelli et al. (1993), who suggest a complex constraint-satisfaction procedure.

Recently, Shepard (1994) accommodated all deviating results in the apparent motion domain into his theory by claiming that we do not necessarily perceive motion in accordance with kinematic geometry whenever the percept is under-specified, as in apparent motion or imagery. Rather, he makes the weaker claim that kinematic geometry is "more deeply internalized than physics" (p. 7). This claim is too weak to be the basis for any predictions. If we take another intuitive physics example, what would Shepard predict for the following case? Imagine a marble that is rolled through a C-shaped tube, which is positioned horizontally on a table top. What will its movement path look like after it exits the tube? If the observer has internalized an approximation to Newtonian mechanics, she should imagine the marble to continue its path in a straight line perpendicular to the tube's opening (Fig. 3, case A). If on the other hand, curved geodesics are internalized, a curvilinear path might be preferred (Fig. 3, case B).[1] Empirically, many subjects erroneously think that the marble should continue to curve, presumably because it has acquired a curvilinear impetus (McCloskey et al. 1980). However, observers who make erroneous predictions prefer the correct straight path when confronted with visual animations of a variety of straight and curved trajectories (Kaiser et al. 1985a; 1992). Thus, only with less visual support are curved paths preferred. Have curved trajectories beyond Chasles' theorem been internalized, or does internalization fail here because it can predict all interesting outcomes?

## 3. Doubts about the epistemological status of internalization

The above examples show that the internalization hypothesis is in trouble. Taken together, those candidates of the internalization hypothesis that are amenable to empirical testing call for a revision of this concept. The literal interpretation of internalization is faced with heavy counter-evidence. The more likely abstract interpretation suffers from two very different problems that have to do with fundamental limits to its empirical verification. The first problem concerns the resolution or generality of the internalized rules, and the second concerns the need for a criterion that determines when internalization has occurred.

### 3.1. The resolution problem

If we say that an organism has internalized a particular regularity or rule, such as the periodicity of the circadian rhythm, we could refer to a very coarse level of resolution: some vague expectancy of day following night. On the other hand, we could mean that the organism has an internal clock and knows down to the minute when the sun will rise. The higher the level of resolution, the easier an empirical test. The level of resolution that we apply to the internalization hypothesis determines to what extent it is amenable to empirical testing.

Kubovy and Epstein (in this issue, p. 621) claim that the internalization hypothesis "has no obvious empirical content and cannot be tested experimentally." This is only true in its broadest reading. In support of Shepard, I not only hold that there are other readings that can be tested experimentally, I also claim that the more fine-grained the operationalization of his hypothesis, the easier it is to refute. For example, the hypothesis that we have internalized the rule that water surfaces at rest are always horizontal is a strong case that allows distinct predictions: we should resolve ambiguous perceptual situations in this context with errors toward a preference for horizontal orientation. On the other hand, the hypothesis that we have internalized some abstraction of this regularity would not necessarily put us in a position to use a few empirical observations of non-horizontal solutions as evidence against the internalization hypothesis. It may not be falsifiable at all if we cannot think of any behavior that could contradict the claim (see Popper 1935).

In Figure 4, I have tried to depict the relationship between postulated internalizations and hypothesis testing. The resolution of a given internalization hypothesis tends to correlate highly with its amenability to empirical testing. Highly resolved statements that claim generality are easy to

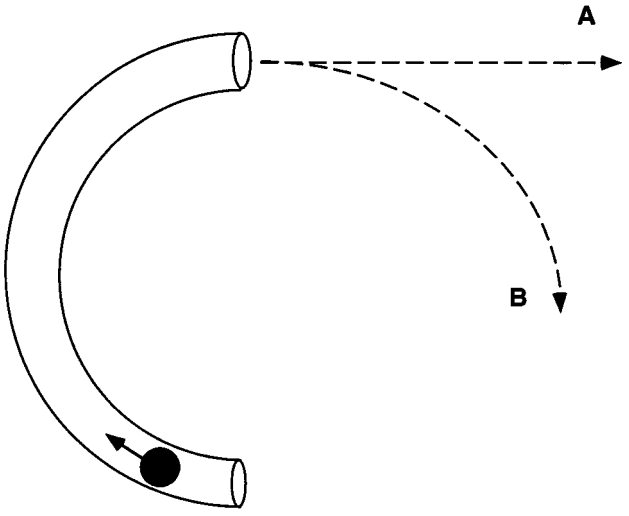



Figure 3. When asked to predict the path taken by a ball rolled through a C-shaped tube, many subjects mistakenly chose a curved trajectory (b) over the correct straight path (a).

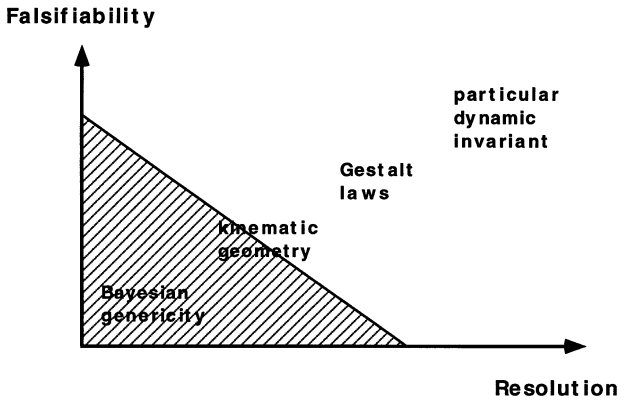



Figure 4. The internalization concept can be analyzed at different levels of resolution. The more specific and the better resolved an instantiation of the claim, the greater its chance to be found false after empirical testing. Ill-resolved claims that are hard or even impossible to falsify are immune to criticism (shaded area).



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

falsify and therefore desirable, as for instance the hypothesis that "all perceived apparent motion trajectories follow geodesic paths." Unfortunately, Shepard's internalization hypothesis is most appealing where it is least resolved. It may in fact be so appealing because it is immune to empirical testing. Also the notion of kinematic geometry and some Gestalt laws are on the brink of immunity as long as they are not supplemented with precise predictions, as for instance the Gestalt law of proximity ("objects in close spatial proximity tend to be perceptually grouped together"). Thus, when discussing *the* internalization hypothesis, we always have to add at what level of resolution we are making our argument. The above distinction of a literal and an abstract interpretation was an attempt to do so.

### 3.2. The criterion problem

The criterion problem refers to the content of the internalized knowledge. What type of knowledge can in principle be internalized? To answer this question, we need to narrow the concept of internalization. In its ill-resolved form, internalization can accommodate such diverse approaches as indirect and direct theories of visual perception. According to the former, without further assumptions, the visual system could not arrive at unique interpretations of the necessarily ambiguous retinal stimulus (Rock 1983; von Helmholtz 1894). Visual perception has to solve ill-posed problems that have no unique solution (Poggio 1990). Assumptions that transform ill-posed problems into well-posed ones are typically not arbitrary, and the percepts they create are not qualified by a question mark (as in the case of a Necker cube, whose percept can change momentarily), but are usually stable and distinct. In other words, the decoding of the stimulus information requires methods of induction (Braunstein 1994) and additional assumptions about the world, which the visual system has – in some broad sense – internalized (Shepard 1984). If this is the case, perceptual problems should not only be solved by the visual system, they should be solved in a manner consistent with laws that govern the physical world. Direct theories of perception (Gibson 1979) would phrase the same basic story rather differently. The makeup of the visual system, as developed through evolution, prepares it to pick up information relevant for proper action. In a sense, internalization is implicit here.

To sharpen the criterion for internalization, a minimal requirement seems to be that the organism must have had a chance to fail to internalize the knowledge in question. Truly universal *a prioris* of perception would thus not be candidates for internalization. Take, for example, the law of noncontradiction: if an object could at the same time exist and not exist, neither object recognition nor epistemology could work. Proponents of evolutionary epistemology argue that our evolutionary world knowledge has no choice but to work with these necessary constraints. This entails that they are also reflected in perceptual processes (Wächtershäuser 1987). This holds not only for laws of logic, but also for basic structural symmetries between the world and perception that may not be coincidental. Campbell (1987), for instance, points out the striking coincidence that almost all objects that reflect or absorb light also block our locomotion and, likewise, all objects that are permeable to light do not obstruct our locomotion. We can see and walk through air, to a lesser degree through water, and not at all through solid objects. This corresponds to the two fundamental constituents of the terrestrial environment, which Gibson (1979) construed to be media and surfaces (of substances). Their existence is too basic to be called "internalized" in any meaningful fashion. Likewise, the optics of the lens, the location of our eyeballs, and so on, impose constraints onto the visual system that need to be considered to understand vision, but do not qualify as examples of internalization.

Internalization also does not need to be an explicit or declarative knowledge structure. It can nontrivially be achieved by virtue of the makeup of the system. This has been conceptualized by direct perceptionists with the use of an analogy. The visual system acts like a smart device (Runeson 1977), as does, for instance, a speedometer. A speedometer does not measure time or distance, thus has no *knowledge* about speed, yet nonetheless "measures" speed by means of an induction current caused by the revolutions of the wheel and translated into the position of the speedometer needle. Taking advantage of induction is, in a manner of speaking, evidence for the fact that some principles of physics have been internalized by the speedometer. Likewise, the visual system can be said to have internalized some world knowledge if its behavior is smart.

In sum, to fulfill the criterion of internalization, a regularity has to be nontrivial and must have a chance to be ignored. Only these cases are subject to empirical testing. A second requirement, not explicitly imposed by Shepard, is best described in terms of Aristotle's classification of causes. If the regularity reflects knowledge of efficient causes in the environment (*causa efficiens*) it can be internalized. If instead the regularity reflects other knowledge, such as action goals (*causa finalis*), it cannot be internalized. Thus, to support the internalization hypothesis, we have to witness the use of rules that are finely resolved and that reflect world knowledge. Since there are many counter-examples at the high resolution end, and a good deal of arbitrariness at the low resolution end, is the internalization hypothesis valuable at all?

## 4. Externalization rather than internalization?

The answer to this question can still be positive if, apart from empirical evidence, we have another method to assess the fruitfulness of the internalization hypothesis. I posit that we do. A thought experiment that supposes an opposite principle may be able to generate important insights and help us decide whether we want to retain the internalization idea. I suggest externalization as the opposite principle. If this opposite principle leads to predictions that are clearly erroneous, we are likely to be on the right track with internalization. If, on the other hand, we apply the cumulative empirical evidence for internalization to the externalization hypothesis and it fares as well or better, then there is something wrong with internalization.

Principles guiding our perception in cases where the percept is under-specified may be a projection of our own body dynamics onto the perceived reality. In other words, *externalized* aspects of the motor system rather than internalized aspects of the physical world outside ourselves may provide defaults for our perceptions. This would require a logic opposite from that of internalization. The logic behind the idea of internalization is that some laws governing the universe have been generalized and incorporated into the vi-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

sual system. An externalization logic, in contrast, would not focus on the receptive visual system but on the active motor system. Considering that the visual system has evolved to guide action and not to give us nice pictures of the world, this appears equally plausible. In a way, the visual system might have "internalized" features of its own motor system. Evidence for this route could be derived from the many instances of ideomotor-action, which demonstrate a very close link between the two systems (see e.g., Prinz 1987). The motor system, in turn, has of course evolved under the constraints that act in its terrestrial environment and therefore exhibits many features that fall under the realm of classical mechanics. The important difference lies in the fact that the motor system is action-oriented and generates its own forces. If default interpretations performed by the visual system are mediated by the action-oriented motor system, a very different set of laws might smarten the visual system. These laws are not abstracted versions of Newton's motion laws but, rather, abstracted versions of force-producing body mechanics.

Thus, let us consider how the visual system turns ill-posed questions into well-posed ones: it derives its default solutions from implicit knowledge of the motor system that it serves rather than from abstract universal laws that have observable consequences. This process might best be referred to as an externalization of body mechanics. An example that illustrates and empirically supports this idea is reported below. It concerns the understanding of ballistic trajectories of projectiles, which can be traced from Aristotle to our times.

### 4.1. Projectile trajectories

Not only our explicit intuitive knowledge about the dynamics of moving objects (Shanon 1976), but also our perceptual knowledge about these events is often erroneous – albeit to a lesser extent (Kaiser et al. 1992). For example, explicit knowledge about trajectories of falling objects is seriously flawed. Similar to the above-mentioned belief that a marble rolled through a C-shaped tube preserves its curvilinear impetus, many people believe that objects dropped from a moving carrier fall straight down, as if they lose their horizontal velocity component (McCloskey et al. 1983). Correspondingly, adult subjects do not favor the correct parabolic trajectory over other paths.

Hecht and Bertamini (2000) presented drawings of various possible and impossible trajectories of a baseball thrown over a large distance. Subjects favored a sine wave path over the parabolic trajectory, and generally all paths of some continuous curvature were judged to be fairly natural. The canonical trajectory shape was neither preferred nor singled out as special. This lack of perceptual understanding, if not evidence for kinematic geometry, might explain why beliefs about the shape of ballistic trajectories were rather warped from Aristotle's times through the middle ages. It was not until the early seventeenth century that Galileo suggested the correct parabolic shape (Wunderlich 1977) that ensues when neglecting air resistance.

In 1572, Paulus Puchner devised an interesting analysis of ballistic trajectories to instruct canoneers of the Saxonian artillery, as visible in Figure 5. Puchner was the weaponry expert at the court of the Saxon elector August. Puchner based his state-of-the-art prediction of cannonball trajectories and distances on the Aristotelian notion of a three-step flight path (Wunderlich 1977): a straight ascension phase followed by a curved arc phase, and a straight vertical drop. This three-step trajectory is not easily compatible with medieval impetus theory, because the circular phase of the flight path cannot be explained by air resistance diminishing the original impetus but only by gravity (or something else) continuously acting on the cannonball. The last straight-down phase was probably an empirical observation that cannonballs tended to drop from almost straight above.

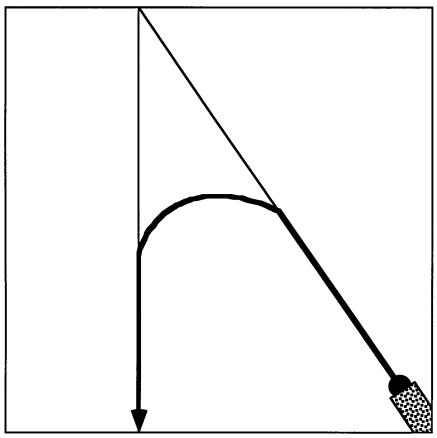


Figure 5. Ballistic trajectories as devised by Paulus Puchner in 1572. His state-of-the-art prediction of cannon ball shot distances was based on the Aristotelian notion of a three-step flight path. First a straight ascension path, second a circular arc, and finally a straight vertical drop. The first and last steps always had to complete a triangle at a given height when extended.

Notwithstanding these conceptual errors, observers have some visual knowledge about the correct parabolic trajectory and even better productive knowledge, as Krist and colleagues (Krist et al. 1993; 1996) have demonstrated. Their moving observers had to hit a stationary target on the ground by dropping an object. Given this facility, adult observers could easily have "internalized" the fact that cannon balls or rocks reach their maximal velocity when they exit the gun barrel or the pitcher's hand. The horizontal velocity component decreases as a function of drag and its change typically remains small in comparison to the deceleration of the vertical velocity component. The latter first decreases to 0, then the ball gains vertical acceleration on the downward part of its trajectory.

Surprisingly, many subjects believe that a ball will continue to accelerate after it has left the pitcher's hand. This belief is mirrored in perceptual judgments when impossible accelerating ball throws were presented in computer animation (Hecht & Bertamini 2000). Figure 6 depicts a trajectory that was judged to be rather natural. The cross indicates the point on the trajectory where observers believed the ball to possess maximal velocity. While these conceptual and perceptual data are grossly incompatible with any law of classical mechanics, including medieval impetus theory, they accurately describe the movement of the pitcher's arm. The latter does accelerate the ball and this accelerating movement might be projected onto the further movement of the projectile. These findings suggest that observers judge the throwing action as a whole and fail to parse the motor action from the mechanical event, or in other words,



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

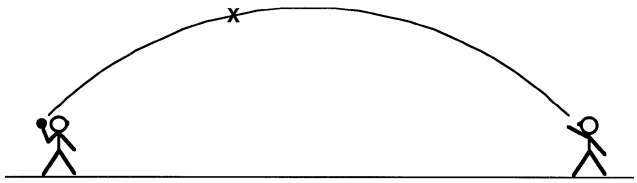


Figure 6. "Consider the path of a ball thrown from a pitcher (on the left) to a catcher (on the right). Mark the point where the ball has maximal speed." When asked this question, subjects' averaged answers correspond the position marked with a cross.

that the body mechanics of throwing have been externalized and projected onto the inanimate projectile.

A different interpretation may consist in the failure to conceive of the ball as an inanimate object. Obviously, as soon as the ball is no longer a projectile but has its own means of propulsion, an indefinite number of acceleration and deceleration patterns are compatible with the laws of physics. This interpretation is, however, rather unlikely given the evidence that even preschoolers are able to distinguish animate from inanimate objects (Massey & Gelman 1988). The dynamic motion context together with the mature age of the observers used in the above examples suggest that the ball was taken to be an inanimate object in these situations (see Gelman et al. 1995; Kaiser et al. 1992).

Even if the notion of externalized body dynamics appears to be rather speculative at this point, the misconceptions as well as the perceptual judgments reflect, in some sense, a continuation of a completed motor action (the acceleration of the arm) and an anticipation of a future motor action (deceleration of the arm and catch of the ball). This piece of evidence, at the least, demonstrates that there are potential competitors for the concept of internalization. If internalization is understood as a principle of abstraction that is prevalent in situations of underspecified perception, it must be legitimate to extend it to the realm of intuitive physics where perception and cognition overlap. In the above example, observers do not behave as if they have internalized regularities about projectile motion. An externalization account fits the data much better. Thus, our final attempt to provide support for internalization by comparing it to its opposite has failed. It even looks as if the notion of externalization has to be taken seriously in its own right.

## 5. Conclusion

I have tried to put the notion of internalization to a test while factoring in as many interpretations of the concept as possible. I have argued that internalization, interpreted in a narrow sense, is false. At the same time, broader interpretations run the risk of making the concept immune to any attempts to test it empirically. The middle ground is what deserves discussion. I have scrutinized this middle ground by recourse to examples from intuitive physics for two reasons. First, they are true to Shepard's spirit of looking for the relation of perception to the laws of physics. Second, they could logically have been internalized. I found mostly problems and counterexamples. Observers do not behave as if they make use of knowledge about gravitational acceleration, or the law of horizontal liquid surfaces. Neither do geometric geodesics prescribe our perceptual solutions in more than a few specialized cases. Internal knowledge about world regularities seems to be specific and task dependent, not universal.

### 5.1. Internalization versus habit

Shepard's concept can be understood as the phylogenetic complement to James's notion of habit. A habit, according to James (1890/1950), is a law that the organism has acquired during its lifetime and it thus has a clear ontogenetic character. Habits necessarily disqualify as internalized because they are acquired and can be changed. At the abstract level of analysis this appears to be acceptable. At the level of concrete examples, however, the distinction between habit and internalized rule is very hard to make. Take the case of a simple reaching movement. In a force field that is not typical for the gravitational field on earth, for instance when being spun on a centrifuge, observers do not take the unexpected forces into account. Their reaches are perturbed. However, after a few more reaches they have adapted to the unusual forces acting on their arms (Lackner & Dizio 1998). Does it make a difference in this case whether we say that the observer has formed a new habit or that she has quickly overcome the internalized regularity? Or has the normal gravitational force field (1 g) only been internalized when people fail to adapt to the new environment? Maybe resistance to adaptation can be used as measure for internalization. By spelling out the differences between habits and internalized regularities in such exemplar cases, the latter concept might be sharpened.

Two aspects of the visual system that are well captured by habit seem to render Shepard's approach cumbersome. First, the visual system is flexible. As in the water-level task, observers may change their behavior dramatically with experience. Experienced waitresses and bartenders produced larger deviations from horizontality than naive participants. Does that mean that internalized regularities can be modified on the spot? If this were the case, it would be almost impossible to differentiate between habits and internalization.

Second, Shepard's model excludes the natural environment as the major player in shaping the percept. The examples that I have discussed here attempted to focus on natural viewing situations. They failed to support his model. A system of bounded rationality (Simon 1969; 1990) such as the visual system may confine its solution space not by resorting exclusively to internal laws but rather by including environmental "satisficing" constraints that come to the fore depending on the environment and the situation in which the actor finds herself. This position could easily be extended to include the body dynamics of the actor as additional constraints.

### 5.2. Can the concept of internalization be salvaged?

Shepard's concept can deal in four ways with the failure to find evidence for regularities such as gravity and horizontality. First, all those potential invariants that did not pass empirical testing could be explicitly excluded from the theory. This would narrow the concept of internalization to a small set of applications.

Second, only certain regularities could be predetermined to qualify for internalization. This solution is also unacceptable. Thus far, Shepard has not provided any rules to decide when a regularity needs to have been internalized. As long


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

as these are missing, the notion of internalized constraints does not have the status of a theory. It cannot be falsified. Shepard's statement could be formulated as: "Some regularities of the physical world have been internalized and act as constraints on perception and imagery." This is an existence statement which can only be disproved if we fail to find a single supporting case. Unfortunately, existence statements by themselves do not allow any predictions about other cases.

Third, another unacceptable salvage attempt would be to push the degree of abstraction of the concept even further. In some abstract sense, it cannot be wrong to claim that we have phylogenetically internalized some regularities of the physical world.

> When formalized at a sufficient level of abstraction, mental principles . . . might be found . . . perhaps attaining the kind of universality, invariance, and formal elegance . . . previously accorded only to the laws of physics and mathematics (Shepard 1994, pp. 2–3).

However, this venue of hyper-abstraction leads to immunity and removes internalization from the empirical discourse.

Finally, the only acceptable solution I see is to make the concept of internalization more powerful by adding specific hypotheses that rule out alternative explanations, such as ontogenetic learning of the circadian rhythm. Given the structure of Shepard's argument, such hypotheses should be derived from evolutionary theory. They might add some of the required resolution to the debate. It remains to be seen whether such salvage attempts are going to be worth the effort.

As an alternative to the exhaustion of salvage attempts, other competing concepts to internalization need to be taken seriously. I have assessed whether support for internalization could be derived from the failure of its opposite: the notion that we have not internalized world knowledge but externalized volitional and motor aspects of our own organisms. This opposite – externalization – fared quite well and merits further exploration.

ACKNOWLEDGMENTS

I wish to thank the Forschungsgruppe "Perception and Evolution" at the ZiF for raising the topic of internalization. I am indebted to William Epstein, Jessica-Gienow Hecht, Mary Kaiser, Michael Kubovy, Donald Hoffman, John Pittenger, Dennis Proffitt, Robert Schwartz, and four anonymous reviewers for valuable comments and suggestions.

NOTE

**1.** I realize that Chasles' theorem does not apply here. Kinematic geometry may not have a clear prediction for this case. However, a general default of curved movement would. And as we have seen, the degree of abstraction appropriate for internalization is highly debatable.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 618–625
*Printed in the United States of America*

# Internalization: A metaphor we can live without

**Michael Kubovy and William Epstein**
*Department of Psychology, The University of Virginia, P.O. Box 400400, Charlottesville, VA 22904–4400*
**kubovy@virginia.edu** **we7v@virginia.edu**
**www.virginia.edu/~mklab**

**Abstract:** Shepard has supposed that the mind is stocked with innate knowledge of the world and that this knowledge figures prominently in the way we see the world. According to him, this internal knowledge is the legacy of a process of internalization; a process of natural selection over the evolutionary history of the species. Shepard has developed his proposal most fully in his analysis of the relation between kinematic geometry and the shape of the motion path in apparent motion displays. We argue that Shepard has made a case for *applying* the principles of kinematic geometry to the perception of motion, but that he has not made the case for *injecting* these principles into the mind of the percipient. We offer a more modest interpretation of his important findings: that kinematic geometry may be a *model* of apparent motion. Inasmuch as our recommended interpretation does not lodge geometry in the mind of the percipient, the motivation of positing internalization, a process that moves kinematic geometry into the mind, is obviated. In our conclusion, we suggest that cognitive psychologists, in their embrace of internal mental universals and internalization may have been seduced by the siren call of metaphor.

**Keywords:** apparent motion; imagery; internalization; inverse projection problem; kinematic geometry; measurement; metaphors of mind

Theorists of perception face two fundamental questions: (1) How does the visual system resolve the inverse projection problem? This problem has been stated as follows: "In classical optics or in computer graphics the basic problem is to determine the images of three-dimensional objects, whereas vision is confronted with the inverse problem of recovering surfaces from images" (Poggio et al. 1985, p. 314). The inverse projection problem is difficult because the environment to stimulation ($E \rightarrow S$) mapping is *noninvertible* (Durbin 1985, Theorem 2.2): for every feature of $S$ there is a corresponding feature in $E$ (i.e., the $E \rightarrow S$ mapping is *onto*), but in all scenes countless different features of $S$ could produce identical features on the retina (i.e., $E \rightarrow S$ is not *one-to-one*). (2) Why is the visual system's solution to the inverse projection problem successful, that is, what accounts for the commonplace fact that the solution typically accords with the way things are?

Shepard responded to these questions with regard to the perception of motion by postulating an internalized kinematic geometry. The most fully developed example of this approach is his treatment of apparent motion. According to Shepard (1994) the fact that apparent movement is perceived at all is owing to the "internalized principle of object conservation" (p. 4). But he is interested in more than the fact that movement is seen. His attention is drawn to the shape of the path adopted by the apparently moving object. Although the number of candidate movement paths joining the two locations is infinite, the perceptual system regularly settles on one. According to Shepard, the preferred path conforms to the principles of kinematic geometry.

In this article, we will claim that Shepard's theory can be divided into two sets of assertions: (1) assertions that the perception of motion is modeled by kinematic geometry, and (2) assertions about the *internalization* of kinematic geometry. We are persuaded by the first set of assertions, and believe that they represent an important advance, but have reservations about the second.

## 1. The inverse projection problem and internalization

Before we turn to Shepard's notion of internalization, a brief review of the history of this idea will provide a useful frame. Our brief excursion will visit Helmholtz, Transactionalism, Rock's cognitive constructivism, Marr's computational theory, and Gibson's ecological approach.

### 1.1. Precursors

Although Helmholtz did not label the problem he was solving as the problem of inverse projection, his theory of per-

MICHAEL KUBOVY is Professor of Psychology at the University of Virginia. He has published empirical and theoretical work on the psychology of visual and auditory perception, with a particular emphasis on Gestalt phenomena and perceptual organization. He has also written about the "pleasures of the mind," the psychology of art, and decision making.

WILLIAM EPSTEIN is Professor of Psychology Emeritus at the University of Wisconsin-Madison and Visiting Professor of Psychology at the University of Virginia. He has published empirical and theoretical work on problems of visual perception.


© 2001 Cambridge University Press
0140-525X/01 $12.50
https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ception is an early attempt to solve it. He proposed (von Helmholtz 1866/1965, p. 153) that perception involves an unconscious deductive inference:

> An astronomer . . . comes to . . . conscious conclusions . . . when he computes the positions of stars in space . . . from the perspective images he has had of them at various times . . . . His conclusions are based on a conscious knowledge of the laws of optics. In the ordinary acts of vision this knowledge of optics is lacking. Still it may be permissible to speak of the psychic acts of ordinary perception as *unconscious conclusions*. . . .

In perception as Helmholtz sees it, one premise of the deductive inference is a law of optics that governs the relation between stimulation ($S$) and the environment ($E$), $S \Rightarrow E$, and is acquired by visual learning guided by tactile experience. The inference follows the *modus ponens* rule of first order predicate calculus: $S \Rightarrow E$, $S$, therefore $E$. For example, the major premise might be a law regarding the way points in the environment give rise to binocular disparity; the minor premise would affirm the occurrent proximal state ($S$), the disparity between the right and the left eye images. The conclusion is an assertion about the world, for example, at a certain position in the environment there is an object of such and such a shape ($E$).

The Transactionalists (Ittelson 1960; Kilpatrick 1961), led by Ames, gave the first clear expression of the inverse projection problem which they labeled the problem of "equivalent configurations." The Transactionalists proposed that the percipient settles on a particular candidate representation by drawing on "assumptions" about the world which assign likelihoods to the candidate solutions.

The Transactionalists' language of assumptions reflects a commitment they shared with Helmholtz which was later promoted by Rock (1983; 1997). It is a commitment to unconscious cognitive processes. According to Rock's cognitive constructivist theory, the perceptual system is stocked with laws (e.g., principles of central projection), and rules (e.g., nonaccidentalness), that serve to direct the solution of the inverse projection problem in an "intelligent" manner. Rock presumes that the knowledge base is represented *in* the perceptual system; it is therefore *internal.*

On the matter of internalization, Rock insisted on distancing his position from Helmholtz's. As noted above, Helmholtz, who was a determined empiricist, postulated that the major premises were themselves the product of individual learning. Rock took a different position. He did not posit that the knowledge base is learned by the individual, but that it is available without prior *individual* learning as the result of learning over the history of the species. Rock proposed to have the cognitive cake without swallowing the indigestible bits of classical empiricism.

The computationalist approach makes the assumption of internalization less urgent. Its goal (Marr 1982; Ullman 1979) is to identify plausible constraints on the environment that can make the $E \rightarrow S$ mapping invertible.

> The main idea for "solving" ill-posed problems, that is for restoring "well-posedness," is to restrict the class of admissible solutions by introducing suitable a priori knowledge. A priori knowledge can be exploited, for example, under the form of either variational principles that impose constraints on the possible solutions or as statistical properties of the solution space. (Poggio et al. 1985, p. 315)

If one can identify the constraints, one has shown how the visual system dissolves the inverse projection problem: When the environment satisfies the posited constraints, the $E \rightarrow S$ mapping is invertible. (It should be noted, as Edelman 1997 has shown, that some important perceptual tasks, such as recognition, may not require inversion.)

It may seem that the computationalists' constraints are merely reincarnations of the Transactionalists' assumptions or of the rules in Rock's (1983) neo-Helmholtzian account. Indeed, some expressions of the computational approach may encourage this interpretation. Nevertheless, it is important to distinguish between the status of assumptions and rules on the one hand, and constraints on the other. For the transactionalists and the cognitive constructivists, assumptions and rules are lodged in the mind of the perceiver. Even though they are not available for conscious assessment they are represented and are causally active in the perceptual process.

Although it makes sense to ask how assumptions and rules are represented and who it is that uses them, these questions are not properly asked about constraints. Computational constraints are *environmental* regularities that have prevailed in the ecological niche of the species over the course of evolution of the perceptual system. As such, they have shaped the design of computational modules so that their output, given optical input under ordinary conditions, is adaptive. The computationalist theorist needs to know the relevant constraints to proceed to the algorithmic level of explanation. This should not be mistaken to mean that the perceiver needs knowledge of the constraints to see the world as she does. The difference between the cognitive constructive stance and the computational stance may be summarized simply: For the cognitive constructivist, the perceptual system *follows* rules; for the computationalists, the system *instantiates* them.

Gibson (1950; 1966; 1979) took a more radical step of denying the need for assumptions, rules, or constraints, therefore making the question of internalization moot. He thought that the so-called inverse projection problem is a pseudoproblem, owing its origin and persistence to a mistakenly narrow construal of both the objects of perception and the nature of stimulation. He argues that the object of perception is a fully furnished world, not objects detached from settings (or isolated from any setting), and that the stimulations that count for a perceptual system are dynamic structures of light and not static slices of spatiotemporal optical structures. Whoever adopts this (ecological) stance, has described the organism's ecology ($E'$) so that the $E' \rightarrow S$ mapping is both one-to-one and onto, and is therefore invertible. This is a world in which the inverse projection problem does not appear.

Although Gibson's explicit position on the inverse projection problem and internalization may appear to be very different from the stance of the computationalists, his implicit position is actually similar to theirs. His assertion that stimulation and the environment are unequivocally linked, or that stimulation carries "information," is in fact tacitly contingent on the satisfaction of environmental constraints. Perhaps the reason that Gibson was reluctant to give this contingency the prominence it later received in the writings of the computationalists was a fear that talk of constraints so readily slips over into talk about mental entities. Had Gibson become convinced that there was a noncognitive formulation of constraints he might have admitted them explicitly into his theory.

Let us summarize what we have said on the issue of the inverse projection problem and internality. According to



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

the approach of Marr and Gibson, constraints are neither (1) lodged in the mind, nor (2) are they active constituents in the perceptual process. They are the conditions which the world must satisfy if the computational algorithms are to go through (Marr 1982) or if the claims for information in the spatiotemporal optical structures (Gibson 1966; 1979) are to be sustained. As such, constraints are passive guarantors or underwriters that are external to the perceptual process (representational transformation for Marr; "pickup" of information for Gibson). We should therefore not view them as internal, and hence not internalized. According to the Transactionalists and Rock, assumptions and knowledge are mental contents that are active in the perceptual process. They hypothesize that assumptions and knowledge direct the process of selecting the best fitting (most likely) distal attribution. It is but a small step from that hypothesis to internalization.

### *1.2. Locating Shepard*

How should we locate Shepard's position in this theoretical landscape? At various times he has aligned himself with Helmholtz's stance (see Shepard 1990b, for example); at other times he has resonated to Gibson's resonance metaphor (Shepard 1984, for example). In the target article he does not signal his position. However, our assessment is that Shepard's position is in the neighborhood staked out by Helmholtz and Rock. Like Helmholtz's premises and Rock's rules, Shepard's universals are deemed by him to be mental contents actively engaged in the perceptual process.

What are Shepard's grounds for positing perceptual universals? In the main, perceptual universals are inferences from behavior: observations concerning the preferred shapes of the paths of movement over many studies of apparent motion lead Shepard to attribute kinematic geometry to the visual system. In this respect, his tactics are similar to the procedures adopted by the Transactionalists in inferring the action of assumptions and by Rock (1983) in his *Logic of Perception.*

Why does Shepard suppose that the universals are *internal*? He provides support for this claim by showing that when we purge all support for a percept from external stimulation, the preferred perceptual solutions conform to the putative universals. The paradigm case for Shepard is the invariant period of the earth's rotation mirrored in the circadian period of activity exhibited by animals maintained in a laboratory environment of invariant illumination and temperature. The analogue in human perception is the apparent motion display. In this case all the normal supports for motion have been removed. When under these circumstances, observers nonetheless perceive motion and the paths of movement they see vary in shape in certain systematic ways, Shepard postulates the action of invisible internal principles.

How do these principles of the geometry of motions become internal? Like Rock, Shepard's answer is to posit a process of "internalization"; a process of gradual acquisition driven by natural selection over evolutionary history.

## 2. Internalization as theory

### *2.1. Critique of internality*

**2.1.1. Kinematic geometry as a model.** We will couch the first part of Shepard's answer in slightly more formal terms than he does. Why? Because we wish to avoid confusion by rigorously maintaining the distinction between models and phenomena, which we have done by thinking of kinematic geometry as a measurement model for the perception of motions.

In everyday language the idea of measurement is so deeply embedded, that we do not make the distinction between the fact that object $a$ is heavier than object $b$ and the fact that object $a$ weighs more than object $b$, that is, numbers are assigned to objects, for example, $\phi(a)$ to $a$, such that if $a$ is deemed in some empirical fashion to be heavier than $b$, then $\phi(a) > \phi(b)$. Measurement theorists have shown (Krantz et al. 1971; Roberts 1979/1984; Scott & Suppes 1958; Suppes & Zinnes 1963) that to understand measurement we must focus on the properties of the numerical assignment. In order to do so, we distinguished (as illustrated in Fig. 1): (1) between empirical objects (e.g., different objects, a, b, . . . ) and mathematical objects (e.g., real numbers, $\phi(a)$, $\phi(b)$, . . . ) And (2) between empirical relations (e.g., "heavier than," $\succ$, and "placed in the same pan of the beam balance," $\oplus$), which apply to physical objects, and mathematical relations (e.g., "larger than," $>$ and "addition," $+$), which apply to elements of the set of real numbers. "Measurement may be regarded as the construction of homomorphisms (scales) from empirical relational structures of interest into numerical relational structures that are useful" (Krantz et al. 1971, pp. 8–9). If we have derived certain fundamental measurement theorems (representation, uniqueness, and meaningfulness) from a set of plausible axioms about the relations that hold among physical objects, then we can guarantee that statements such as the following are true:

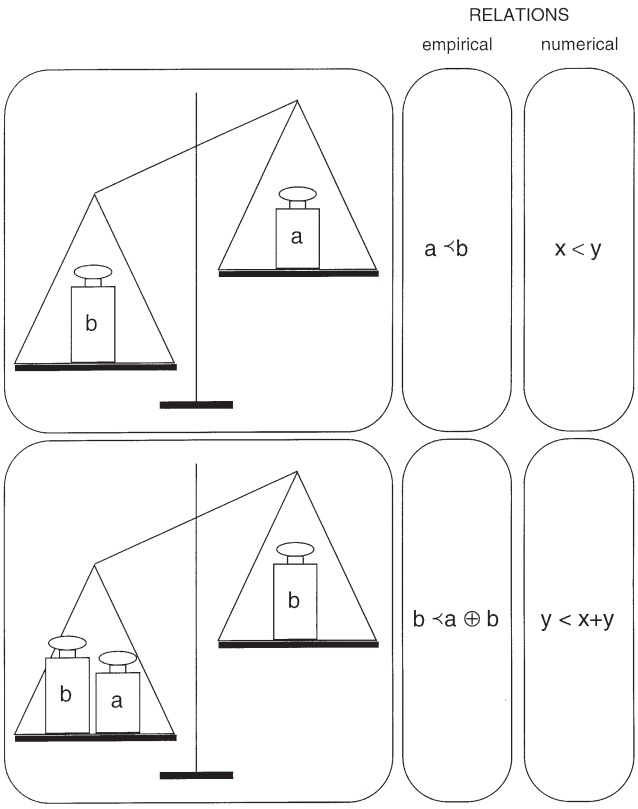



Figure 1. The distinction between empirical objects and relations and mathematical objects and relations.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

> Given an empirical relation $\succ$ on a set $A$ of physical objects and a numerical relation $>$ on the real numbers $\mathfrak{R}$, a function $\phi$ from $A = \{a, b, \ldots\}$ into $\mathfrak{R}$ takes $\succ$ into $>$ provided the elements $a, b, \ldots$ stand in relation $\succ$ if and only if the corresponding numbers $\phi(a), \phi(b), \ldots$ stand in relation $>$. Furthermore, the function $\phi$ takes $\oplus$ into $+$: $\phi(a \oplus b) = \phi(a) + \phi(b)$.

Kinematic geometry is considered to be a branch of mechanics. What does this mean? It means that we can construct a homomorphism from an empirical relational structure of physical motions into a geometrical relational structure called kinematic geometry. It is somewhat more complicated to say what it means when we say that kinematic geometry is a model for the perception of motion. It requires us to substantiate four claims (Fig. 2):

1. There is a homomorphism $k$ from an empirical relational structure of *physical motions* into a geometrical relational structure called *kinematic geometry;*

2. There is a homomorphism $p$ from an empirical relational structure of *physical motions* into an empirical relational structure of *perceived motions;*

3. There is a homomorphism $q$ from an empirical relational structure of *perceived motions* into a geometrical relational structure which is a *model of the perception of motions;* and

4. There is a homomorphism $m$ from *model of the perception of motions* into a geometrical relational structure, *kinematic geometry.*

There are two ways in which conformity between perceived motion paths and kinematic geometry may be construed. One might give the claim a more modest reading and say that kinematic geometry is a model of the perception of motions, as summarized by the list of four homomorphisms just listed.[1] This means that the visual system proceeds *as if* it possessed knowledge of kinematic geometry. We do not think that Shepard wants to be read this way. We think that he wants to persuade us that kinematic geometry is (1) internal, and (2) has been internalized. This is where our disagreement begins.

### 2.2. Questioning internalization

Shepard believes that by removing the percipient from ordinary contact with stimulation we can show that kinematic geometry is internal. With the exception of J. J. Gibson and the ecological realists, there is agreement among students of perception that this is a useful, perhaps indispensable strategy for making the invisible visible. Despite this consensus, there is a problem in *exclusive* reliance on nonrepresentative settings. Consider apparent motion. The apparent motion experimental setting – caveats aside, Shepard (1994, p. 10), for example – is unlike the typical conditions of real motion. So while presumably the special conditions of apparent motion may disclose the operation of hidden principles, they cannot – taken alone – establish that these hidden principles are implicated in perception of real motion.[2] Most investigators who adopt the tactic argue that what is uncovered under the special conditions applies generally, including also the representative conditions. But we should sign on to this petition only when there is evidence to support it. This requirement can, in fact, be satisfied in certain cases. However, as matters stand currently, there is no evidence in Shepard's work that shows convincingly the involvement of *internal* principles of kinematic geometry in the determination of perceived real motion.[3] And, in places Shepard seems to suggest that he considers the geometry-compliant solutions to be default solutions that should not be expected to appear under conditions of ordinary motion perception: " . . . (under favorable viewing conditions) we generally perceive the transformation that an external object is actually undergoing in the external world, however simple or complex, rigid or nonrigid. Here, however, I am concerned with the default motions that are internally represented under unfavorable conditions that provide no information about motion . . . " (Shepard 1994, p. 7).

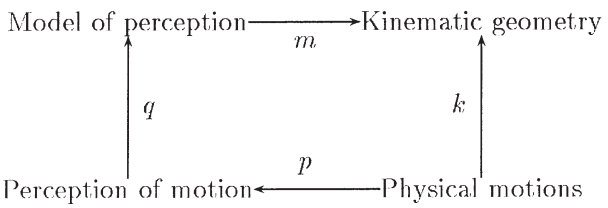



Figure 2. The simplest construal of the relation between kinematic geometry and perception.

It is not incoherent to hold that independent and different principles govern perceived motion paths under such radically different circumstances as those which prevail in the cases of apparent motion and real motion. However, in the context of Shepard's internalization hypothesis an apparent paradox arises. If it is supposed that in the case of ordinary motion – which served up the grist for the evolutionary mill and a process of internalization – there is no evidence of a role for geometrical constraints then how could our distant ancestors have internalized kinematic geometry for application to the special case of apparent motion? It is easy to imagine generalization from the ordinary case to the special case, assuming counterfactually that ordinary real motion did exhibit a determining role for geometric constraints. But it is hard to see how internalization of kinematic geometry could have proceeded independently of the same development in the case of ordinary motion.

Even when the preceding concerns are set aside, the internalization hypothesis suffers from a number of shortcomings. It has no obvious empirical content and cannot be tested experimentally. Moreover, the cash value of the "internalization" hypothesis is questionable. Functionalism, as a starting point, is not what is at stake here. We agree that questions of function (what Marr 1982 has called the "computational theory") are very important. It matters whether one supposes that the function of the heart is to pump blood or to produce audible thumps. And it matters whether one supposes that the function of the perceptual system is to deliver descriptions (representations) of the environment (Marr 1982) or to support action (Gibson 1979) or both (Milner & Goodale 1995). While an argument can be made that neither functionalism as an -ism nor evolution by natural selection is necessary in assessments of function, we will accept that an evolutionary stance helps focus attention on function. (This stipulation notwithstanding, in the history of biology few discoveries rival Harvey's discovery of the function of the heart which Harvey made without the help of evolutionary theorizing or a self-conscious functionalist stance.)

The matter at issue here is, in what way does speculation concerning the origins of a function in the remote part (Pleistocene) contribute to an understanding of the process that subserves the function?



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Is there a single parade case that can be trotted out to show the power of the internalization hypothesis to reduce the number of candidate process hypotheses? We cannot think of a case in the field of perception. Consider two hypothetical knowledge states: (1) In the first state, we have come to know in great detail the brain structures and activity underlying stereopsis by studying the contemporary visual system exclusively, with no reference to phylogenetic development. (2) In the second state, we know everything that we know in the first state and in addition – based on a study of the fossil record – we also have come to know (that is, to tell a plausible story about) the evolution of the brain structures. What is the advantage of 1 over 2 when it comes to understanding how the visual system computes depth from disparity? We do not see how the knowledge of origins constrains the choice among the plausible algorithms.

In places, Shepard suggests that the internalization route to acquisition avoids the embarrassments of the rival origins story which "leaves it to each individual to acquire such facts by trial and possibly fatal error" (1994, p. 2). In this respect, Shepard and Rock are moved to posit internalization by the same considerations. But the argument from selective refutation can be misleading when it causes us to fail to notice that the surviving hypothesis has its own defects. A condition that must be satisfied for plausible postulation of internalization is that an enduring pervasive external regularity must be obvious. The exemplary case cited by Shepard is the external day-night cycle and the internal sleep-wake cycle. But where is the pervasive external regularity in the case of object motions? As Proffitt[4] reminds us, "real motion cannot violate kinematic geometry" and therefore the laws of kinematic geometry are a superset of the laws that could be extracted by the visual system from the data offered by real motions. (*D* in Fig. 3). Therefore, only the latter might have been internalized. As Todorovič (1996) has remarked:

> A source of the problem may be an inappropriate analogy between the operation of the perceptual system and the operation of scientific inquiry. Concretely, the idea is that the principles that the perceptual system has extracted from the environment during evolution are in certain aspects analogous to the principles, such as Chasles's Theorem, that the scientific community has formulated in the last centuries. Such an idea is intriguing, but it should be treated with caution. Perceptual and intellectual activities are to a certain extent related, but they are also quite different. (pp. 17–18)

Because, as we have just seen, the laws of kinematic geometry are a superset of the laws of real motions, Proffitt and Kaiser (1998, sect. II.D, pp. 184–90) see the relation between internalization and evolution differently. They point out that Cassirer's (1944) view of perception is a powerful argument against the belief that the conformity of perception to kinematic geometry has its roots in ontogeny, in keeping with Helmholtz's thought, or in phylogeny, in keeping with Rock's ideas. In order to make sense of Shepard's views, Proffitt and Kaiser (1998) read him as if he were following Cassirer's neo-Kantian convictions. We think that too much of what Shepard (1994) says would have to be ignored to interpret him this way.

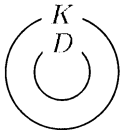



Figure 3. The relation between the laws of kinematic geometry (*K*) and the laws that can be extracted from data provided by real motions (*D*).

## 3. The pragmatics of internalization

Having read our argument against the use of the concept of internalization, the reader may ask, why are the authors trying to legislate the use of the term internalization? After all, science does not progress through a progression of terminological refinements, but through a cycle of progressively more refined data collection and theory construction.

We have chosen to answer this question in four ways. First, with a distinction between two kinds of terminological strictures in psychology. Second, with a case study: how an apparently innocuous choice of language, by Shepard, led to subsequent confusion. Third, with a comparison of Shepard's use of internalization with the use of cognate terms used in psycholinguistics. Fourth, with a discussion of the appeal of metaphors in the formulation of theories.

### 3.1. Two kinds of terminological strictures

We would like to make clear that our goal is to persuade our colleagues to take particular care in the choice of *theoretical* terms, which is different from trying to reform the use of *descriptive* terms. An example of the latter occurs in the literature on anthropomorphism. For example, Crist (1999, p. 22) quotes Barnett (1958, p. 210):

> Darwin . . . took it for granted that terms like *love, fear,* and *desire* can usefully be employed to describe the behaviour of animals – or at least of mammals – generally. He accepted the colloquial use of the word emotion. In doing so he assumed (by implication) that other species have feelings like our own. . . . Since his time it has gradually been found more convenient to describe animal behaviour, not in terms of feelings of which we are directly aware only in ourselves, but in terms of the activities which can be seen and recorded by any observer; we may also try to describe the internal processes that bring these activities about.

Barnett does not propose to purge "emotion" from the *theoretical* vocabulary of psychology:

> If the word emotion were to be used in the scientific study of animal behaviour, its meaning would have to be shifted from the familiar, subjective one: it would have to be used to refer, not to feelings, but to internal changes which could be studied physiologically. (Barnett 1958, p. 210)

In other words, if the term was found to be useful in theories about humans and other animals, there would be no harm in suggesting that other species may have subjective experiences similar to those reported by humans.

We are arguing that the use of the term "internalization" in *theories* may lead to confusion. We turn now to an example of how such confusion could arise.

### 3.2. Do scientific terms matter?

One of us (Kubovy 1983), in an enthusiastic review of Shepard and Cooper (1982), had one criticism of this book. It was not a criticism of the methodology or of the theory, but of the way the results were formulated. The authors of the book often used expressions such as "the rotation of men-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tal images." Kubovy thought that there was a danger that such expressions might mislead readers into thinking that more was being claimed about mental imagery than was implied by the data. Where was this rotation taking place? In the mind? In the brain? Kubovy suspected that locutions such as "mental rotation" could lead people to think that a black box had been pried open without so much as an EEG.

Without doubt, the brilliance of the research done by Shepard, Cooper, and their collaborators warranted its enthusiastic reception by a cognitive community that was beginning to mature and was looking for a rallying point. It was a community looking for a phenomenon that would convince psychologists and nonpsychologists alike that the cognitive approach was producing clear and important results. And so the cognitive community was a bit quick to attribute achievements to this research that went beyond the evidence given.

The problem Kubovy saw was created by the way Shepard and Cooper summarized their results: they spoke as if they could show that in their experiments a physical action ("rotation") was being applied to a mental entity ("mental image") whose content was a physical object ("random polygon"). This linguistic form surely implies that something physical is happening to the mental image. Now this sort of careless expression is not uncommon in psychology. For example, we may talk about the decay of memory, which, however subtly, suggests that something (the engram?) is decaying somewhere (in the brain?). Nevertheless, there was less opportunity for readers to be misled when they were learning about memory than when the topic was mental imagery. The reason is this: In discussing imagery, Shepard was also talking about the form of processing that was taking place: it was an analog process rather than a digital one.

We believe that the cognitive community did not only infer from these forms of expression that somewhere an image was rotating, but that the data could support a claim about the *computational implementation* of this rotation: that it was analog rather than digital. Of course this led to a great deal of controversy (Anderson 1978), but the damage was done. The controversy focused on the indeterminacy of the computational implementation, but this focus did not remove the impression that the topic under discussion was the rotation of mental images.

Can one formulate the results in this field without risking being misunderstood? According to Kubovy (1983) one should avoid locutions of the form

(1) * Φ-ACTION(Ψ-CONTAINER(Φ OBJECT)),

where Φ stands for "physical," and Ψ stands for "mental." Instead, we should use expressions of the form

(2) Ψ-ACT(Φ-TRANSFORMATION(Φ OBJECT)).

For example, instead of talking about a person "rotating a mental image of an object," one might say the she is "imagining the rotation of an object." Not only does form (2) avoid sandwiching the mental between two physical descriptions, but it also avoids implying that a mental entity is a thing by turning it into a verb ("imagine") rather than a noun phrase ("mental image").

As an exercise, let us apply this improved form of speech to the question of the analog nature of mental rotation. If instead of wondering whether the rotation of the mental image is analog or digital, we can ask: does the evidence support the claim that when people imagine the rotation of an object, they imagine the object undergoing a continuous rotation? When rephrased in this way, the question of the analog nature of mental transformations seems less of a puzzle, but also less of an achievement, because phenomenology seems to provide a prima facie answer, yes. And furthermore, the data in no way contradict phenomenology.

By the time Kubovy invited Shepard to contribute to the section he edited of the *Handbook of Perception and Human Performance* (Fink & Shepard 1986), Shepard had accepted the criticism and was eager to have the terminology of the chapter conform with Kubovy's terminological suggestion. Kubovy was concerned that this stylistic constraint would lead to awkwardness of expression. We do not think it did, but we invite our readers to judge for themselves.

The topic went dark for a while until Nelson Goodman (1990) expressed concern about cognitive psychologists' talk about "pictures in the mind." In his reply Shepard (1990c, p. 370), cited Kubovy's review and concluded:

> Moreover, the latter, more carful reformulation brings out the essential feature of mental imagery as I recommend we understand it: mental imagery is of external objects, and is therefore to be defined and studied not as some strange, non-material "picture in the head" but in relation to potential test stimuli that are both external and physical.

Unfortunately, it appears that the confusion has not abated. In a dialog between two distinguished French scientists, Jean-Pierre Changeux – a neurobiologist – and Alain Cannes – a mathematician, the latter says:

> Reading your book *Neuronal Man,* I was surprised to realize how much is understood about the brain. . . . I was impressed too by Shepard's mental rotation experiments, in which a subject is asked if two objects are the same after rotating them in three-dimensional space. They show that the response time is proportional to the angle of rotation, and thus that cerebral function obeys physical laws. (Changeux & Cannes 1995, p.5)

### 3.3. What has Shepard been trying to do?

When one is dealing with a figure as important to the history of our field as is Shepard, one sometimes better understands the breadth and depth of the person's thinking by elucidating certain parallels that guide his thinking. We suspect that Shepard has been looking for a model of perception that shares some important features with Chomskyian linguistics.[5] In 1981, Shepard offered the diagram shown in Figure 4 to illustrate his concept of psychophysical complementarity and illustrate its application to mental rotation (Shepard 1981b). It is particularly interesting to note that he calls the internal representation "Deep Structure."

The Poggio et al. (1985) argument that the environment to stimulation ($E \rightarrow S$) mapping is noninvertible is parallel to a similar insight of psycholinguistics: "A device capable of [developing the competence of a native speaker] would have to include a device that accepted a sample of grammatical utterances as its input . . . and . . . would produce a grammar of the language . . . as its output. . . . To imagine that an adequate grammar could be selected from the infinitude of conceivable alternatives by some process of pure induction on a finite corpus of utterances is to misjudge completely the magnitude of the problem" (Chomsky & Miller 1963, pp. 276–77).

More specifically (Pinker 1984), suppose children learned



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

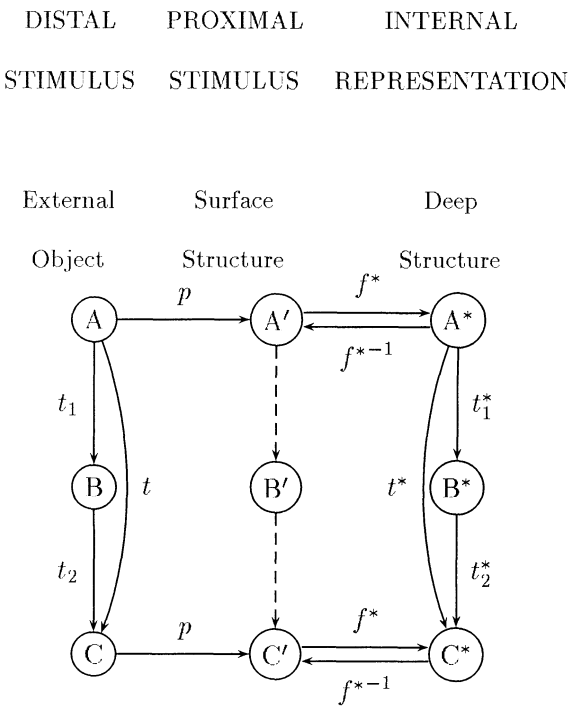



Figure 4. Shepard's schema of the projective ($p$), formative ($f^*$), and transformations mappings between objects (A, B, and C), proximal stimuli (A′, B′, and C′), and internal representations (A*, B*, and C*). (Redrawn from Shepard 1981b, Fig. 10–1, p. 295.)

their language (which we will denote *T* for "target") by induction. Before the child has been exposed to *T*, it could hypothesize what the rules of the language might be (let us denote the set of hypothesized rules by *H*). At that point there will be no overlap between *T* and *H* (Fig. 5a): none of the child's utterances is grammatically well formed. For example, the child never says *We went*, or *We broke it*, and always says *We goed* and *We breaked it*. Because the items of *T* (marked "+") to which the child is exposed can serve as *positive evidence* that *H* is incorrect (because these items are not members of *H*), *H* will grow and come to overlap with *T* (Fig. 5b): The child might say *We went*, but persist

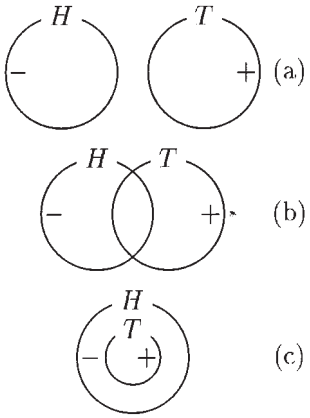



Figure 5. Three situations a child could be in while learning a language. Each disk represents the set of sentences constituting a language. *H* stands for "hypothesized language," *T* stands for "target language."

in saying *We breaked it*. But now the child faces two problems: learning the further rules of *T*, and eliminating the incorrect hypotheses of *H* (marked "−"). This could occur only if parents corrected their children's incorrect grammar. But, according to Pinker (1984), parents do not provide the required *negative evidence*. If this were the case, children would end up knowing a superset of *T* (Fig. 5c). This is one reason why Pinker argues that children are endowed with inborn constraints about the possible form of linguistic rules.

Shepard's argument about the internalization of kinematic geometry is parallel to the linguistic argument. This linguistic argument runs: children would not be able to learn a grammar unless they were endowed with inborn linguistic constraints. These constraints are internalized. Analogously, since kinematic geometry is a superset of what can be observed in real motions (Fig. 3), human beings could not have acquired it unless they were endowed with inborn geometric constraints that corresponded to kinematic geometry. Therefore, the argument runs, kinematic geometry is internalized. The difference between the two domains is this: In language acquisition, the inborn constraints insure that eventually the size of *H* will diminish and coincide with *T*. In perception, the mind comes equipped with *K*, and loses nothing by using kinematic geometry to resolve ambiguities that result from missing information in the sensory input.

### 3.4. Metaphors of mind

Confusing forms of expression with respect to topics in cognitive science could be due to the metaphorical nature of many abstract concepts, as Lakoff and Johnson (1990) showed persuasively in their book *Metaphors We Live By*. Lakoff and Johnson (1999, pp. 235–36) summarize the evidence that "there is an extensive subsystem of metaphors for mind in which the mind is conceptualized as a body." When the MIND IS A BODY metaphor is applied to thinking, two metaphors of smaller scope come into play: THINKING IS PHYSICAL FUNCTIONING and IDEAS ARE ENTITIES WITH AN INDEPENDENT EXISTENCE. When the two metaphors are combined, we get THINKING OF AN IDEA IS FUNCTIONING PHYSICALLY WITH RESPECT TO AN INDEPENDENT EXISTING ENTITY.

In particular, confusion with respect to mental rotation may be attributed to a common form of the metaphor that equates thinking with functioning physically and ideas with independent existing entities. This is the metaphor THINKING IS OBJECT MANIPULATION: "ideas are objects that you can *play with, toss around,* or *turn over in your mind*" (Lakoff & Johnson 1999, p. 240). So it is natural to speak of rotating a mental image. And because more often than not, the metaphorical nature of our thinking is unconscious, it is not easy for the scientist to see how such expressions could be misleading.

The metaphor of internalization is rooted in another member of the same family of metaphors: ACQUIRING IDEAS IS EATING (Lakoff & Johnson 1999, pp. 241–43). This metaphor leads us to compare an interest in ideas to an appetite (e.g., *thirst for knowledge*). We refuse to *swallow* bad ideas, but we *can really bite into* ideas that are *meaty*. The term internalization is but a slightly disguised synonym of ingestion.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

## 4. Assessment and recommendation

Although the notion of internalization is an appealing metaphor, it does not add to the power of Shepard's theory according to which certain important aspects of perception are captured by kinematic geometry. This and other episodes in the history of our field lead us to recommend that we strip our scientific writing of metaphors we can live without, and, until theoretical and empirical progress suggest and support metaphorical terms, that we formulate our theories in as neutral a language as we can.

ACKNOWLEDGMENTS

We wish to thank Arthur Glenberg, R. Duncan Luce, Dennis R. Proffitt, and Sheena Rogers for their helpful comments. This work was supported by NEI Grant 9 R01 EY 12926-06 (Kubovy, PI). Request reprints from either author at the Department of Psychology, Gilmer Hall, The University of Virgina, Charlottesville, VA 22903-2477 or by email: kubovy@virginia.edu or we7v@virginia.edu.

NOTES

**1.** The reader will surely agree that to demonstrate that kinematic geometry is a model of perception in this sense would be a major achievement.

**2.** We hasten to note that there actually is considerable evidence to support the contention that the same mechanisms underlie real and apparent motion. Our argument here is about scientific strategy, not about apparent motion.

**3.** Writing elsewhere on the evolution of principles of the mind, Shepard (1987a, p. 266) remarks that "the internalized constraints that embody our knowledge of the enduring regularities of the world are likely to be most successfully engaged by contexts that most fully resemble the natural conditions under which our perceptual/representational systems evolved."

**4.** In comments on a draft of this article, personal communication, March 10, 1999.

**5.** This section may appear to some readers as an exercise in hermeneutics, but since we were enlightened by it, we thought that our colleagues might be too, and therefore might tolerate our heretical use of a method that would be listed in the *Index of Forbidden Methods* if such an index were prepared.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 626–628
*Printed in the United States of America*

# Evolutionary internalized regularities

**Robert Schwartz**
*Department of Philosophy, University of Wisconsin-Milwaukee, Milwaukee, WI 53201*
**schwartz@uwm.edu**

**Abstract:** Roger Shepard's proposals and supporting experiments concerning evolutionary internalized regularities have been very influential in the study of vision and in other areas of psychology and cognitive science. This paper examines issues concerning the need, nature, explanatory role, and justification for postulating such internalized constraints. In particular, I seek further clarification from Shepard on how best to understand his claim that principles of kinematic geometry underlie phenomena of motion perception.

My primary focus is on the ecological validity of Shepard's kinematic constraint in the context of ordinary motion perception. First, I explore the analogy Shepard draws between internalized circadian rhythms and the supposed internalization of kinematic geometry. Next, questions are raised about how to interpret and justify applying results from his own and others' experimental studies of apparent motion to more everyday cases of motion perception in richer environments. Finally, some difficulties with Shepard's account of the evolutionary development of his kinematic constraint are considered.

**Keywords:** apparent motion; circadian rhythms; constraints; ecological validity; evolution; internalizated regularities; kinematic principle

## Introduction

Talk of evolutionary internalized regularities in perception, although much in vogue, can be vague. One way to sharpen discussion of the topic is to focus on a particular proposal. Roger Shepard's seminal 1984 paper and decade later update (1994; reprinted in this volume) are surely worthy of such attention. Shepard skillfully probes the issues in breadth and in depth. And his ideas have had a major impact, not only in the study of vision but in other areas of psychology and cognitive science. Still, I am not sure I fully understand Shepard's claims in these papers and other elaborations (Carlton & Shepard 1990a; 1990b; McBeath & Shepard 1989). Thus, my discussion may be more fruitfully viewed as an exploration of the issues and a request for additional clarification, rather than as a criticism of Shepard's position.

I begin by briefly exploring general aspects of the nature and notion of an "internalized regularity." Next, I consider Shepard's kinematic principle, questioning the analogy Shepard draws between the internalized circadian rhythms of animals and his proposed perceptual constraint. Problems, then, are raised about the ecological validity of this constraint and the role it might play in the perception of ordinary, everyday motion. In turn, consideration of these issues would seem to pose some difficulties for Shepard's evolutionary account of the kinematic principle.

## 1. Constraints and internalization

To provide a framework in accord with Shepard's own ideas, I think it would be helpful to make explicit the relationship between a constraint and its possible internalization. In Shepard's sense, the claim that a constraint is internal*ized* goes beyond the claim that the constraint is presently internal or is somehow internally represented and functioning. First, the constraint must be inherited or "innate," and not the result of learning. Second, the constraint must come about by a *particular* evolutionary route. Internalized constraints result from the incorporation of features or universal regularities of the external world. If a constraint did not develop in response to a corresponding external regularity, but, for example, only tagged along on the back of another mutation or was a derived manifestation of the interaction of several independently selected evolutionary constraints, it would not, I take it, be counted as internal*ized*.

Members of the species who display the influences of an internalized principle do not themselves do the incorporating. The process of internalization takes place in prior generations as an evolutionary reflection of the environment. Shepard's focus on internalized constraints seems driven in part by the idea that such principles convey an evolutionary advantage. Establishing the specific benefits of an inherited constraint, however, is not an a priori matter. In light of issues discussed below, I am not sure what advantage Shepard's kinematic principle is supposed to confer. Nor am I very clear how and why he thinks the constraint would have come to be incorporated.

## 2. A paradigm case

Shepard offers an example of the circadian rhythms of certain animals as a model for his proposal about human per-

ROBERT SCHWARTZ, professor of Philosophy at the University of Wisconsin-Milwaukee, has written on issues concerning representation, language acquisition, mathematical cognition, and perception. He is the author of *Vision: Variations on Some Berkeleian Themes* (Blackwell Publishers).


© 2001 Cambridge University Press
0140-525X/01 $12.50
https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ception. He points out that although the biological clocks of these animals appear attuned to the environment, the observed behavioral correlations are misleading. The rhythms are not under the control of external stimuli. When the animals are put in artificially altered environments their biological clocks remain largely unchanged. For Shepard the mechanism of circadian rhythms is a paradigm case of an internalized evolutionary constraint whose existence is demonstrated and manifested in its lack of dependence on the immediate environmental situation. The analogy, however, between this paradigm case and constraints on vision, requires further examination.

What is supposedly striking about the biological rhythms of these animals is their relative insensitivity to alterations in environmental conditions. The pattern of behavior continues in spite of relevant changes in the stimuli. But I am not convinced this feature of the paradigm fits all that well with the way some prominent constraints in vision theory are thought to work. Consider, for example, one of the most widely cited and accepted perceptual constraints, the rigidity principle. The visual system, it is maintained, prefers rigid interpretations over nonrigid ones. Yet, perception of objects as rigid does not run off as independently of the external stimuli as the circadian rhythms are said to do. Under normal viewing conditions, a real object that deforms its shape will generally be perceived as such. Where do things stand with regard to the force and function of Shepard's kinematic constraint?

## 3. The kinematic constraint and ecological validity

According to Shepard, his constraint entails that in perceiving motion "one tends to experience that unique, minimum twisting motion prescribed by kinematic geometry" (1984, p. 425). Here, I wish to examine the issue of sensitivity to environmental input raised in the previous paragraph. In particular, how are we to understand the claim that we have a *tendency* to see motion in terms of Shepard's principle of kinematic geometry?

Under normal viewing conditions, if presented with a real object moving along a path that is not the "unique, simplest rigid motion," it is most often perceived veridically. Shepard's kinematic constraint, like the rigidity constraint, then, does not cause or force perceptions that are decoupled from the actual environmental stimuli. Shepard allows as well that even in cases of apparent motion, perceptual experience may not adhere to the constraint. Thus, Shepard must square the fact that we readily see movement in violation of the kinematic principle with the claim that evolution leads us to see the world along the lines of the constraint. His solution to this problem is to claim that failures to satisfy his proffered principle occur as the result of conflicts with other constraints and stimulus conditions.

Shepard attempts to support his theory mainly by appeal to phenomena of apparent motion (and to a lesser extent imagery), not by studies of real object motion.[1] Reliance on this evidence has its difficulties:

1. In emphasizing the importance of biology, evolution, and Gibsonian theory, Shepard is anxious to champion the idea of ecological validity. Now one thing which seems clear is that the conditions and stimuli used in the apparent motion experiments are not especially typical of normal movement perception. Hence, there is the worry that results found under these limited circumstances are not ecologically valid. They may not transfer or apply to cases of real motion in more ordinary environments.[2]

2. I believe Shepard does not deal adequately with this issue, that is, with the possibility that apparent motion studies do not support substantive claims about the role kinematic geometry actually plays in normal perception. Shepard himself notes that constraints will be violated when an alternative interpretation is "forced on the observer by external conditions" (1984, p. 430). But if all it takes to *force* such perceptions on an observer are more or less ecologically standard conditions, the explanatory significance of the supposed internalized regularity is put in jeopardy.

I think Shepard slights this problem, because he wishes to stress the parallels with the circadian rhythms paradigm. Indeed, one of the major methodological lessons Shepard draws from these animal studies is that uncovering evolutionary constraints requires the use of abnormal experimental conditions. His reason is that if a constraint does embody a regularity occurring in the environment, it will remain hidden in ordinary circumstances. For it will seem as though the behavior is simply being caused by instances of that very environmental regularity. To discover constraints on circadian rhythms it was necessary to remove the animals from their ordinary environment and place them in artificially created settings.

Unfortunately, the need to appeal to relatively non-normal conditions is in tension with a commitment to ecological validity. Some of the difficulties surface when one examines Shepard's attempts to account for constraint violations in apparent motion.

## 4. Constraint violations

Within certain temporal and spatial limits, when a circle is flashed on the left and a square on the right, subjects report they see an object go through geometrical shape transformations while traversing the gap. They do not see it as movement of a rigid body. More complicated compressions, expansions, shape changes, along with violations of the unique kinematic path constraint are experienced in numerous other apparent motion experiments.

Shepard is well aware of such findings. His reply is that the rigidity and kinematic constraints do hold, but only under "conducive conditions" (1984, p. 430; 1994, p. 7). Notably, constraints will be violated when, as with an alternating circle and square, the demands of the principles are not consistent with or are in conflict with the stimuli. Processing limitations are said to be responsible for still other apparent motion violations of constraints. For example, Shepard argues that the time from the onset of one stimulus to the onset of the other can be insufficient to allow for the kind of motion required by the internalized principles. An appropriate rigid kinematic trajectory may be too lengthy a path to travel for it to be completed in the time available between the onsets of the two stimuli. Accordingly, the visual system resolves the conflict by "taking" a shorter path. It perceives a constraint violating shortcut path that can be traversed within the given time span. Evaluating Shepard's explanation of these apparent motion phenomena would require detailed examination, not to be undertaken here.

In any case, I do not believe Shepard's account of con-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

straint violations in apparent motion speaks adequately to concerns about the ecological validity of his kinematic principle in more richly structured environments. A claim of ecological validity would perhaps be more convincing if satisfaction and violation of the constraint were to function in ordinary motion perception as it does in apparent motion. But the case for this claim is not so obvious. Generally, a real object moving along a constraint satisfying kinematic path is not perceived as taking a constraint violating shortcut, even when the time duration would provoke a constraint violating apparent motion trajectory. Similarly, real objects moving along constraint violating non-unique paths are generally perceived as such, even when their transit times are of sufficient duration to trigger constraint satisfying paths in apparent motion.

Shepard, then, allows that in many situations the paths and deformations experienced during apparent motion violate supposedly internalized principles. He attempts to resolve this difficulty by explaining away the violations. In order to do this, he offers a set of additional conditions that must be met if apparent motion is to conform to his kinematic constraint. I do not think, however, that the function, effect, and relevance of comparable restrictions have been shown or can be assumed to hold in the perception of more everyday cases of real movement.

## 5. External forces

Establishing a significant role for Shepard's internalized kinematic principle to play in the perception of richly structured, everyday environments remains problematic.

1. When actual motion accords with Shepard's kinematic constraint, the influence of an internalized regularity may be minimal or nil, since as he admits, there may be enough information in the stimulus to "force" the correct perception without its aid. Alternatively, when in everyday circumstances real motion does not fit the countenanced pattern, it will usually be perceived veridically. Once again, the stimulus will be sufficient to force the correct perception. Shepard would seem to need, then, evidence indicating that his constraint continues to function in ordinary environments, environments where the external conditions appear "rich enough" on their own to determine the perception. In Shiffrar and Shepard (1991), subjects' path matching judgements of real movements are taken to support such a claim. Also, perturbation studies might be devised to show the constraint does have influence, or at least has to be "overcome," in perceiving real motions that violate the principle. Were this so, the kinematic principle might be construed along the lines of a probabilistic "soft" constraint – a constraint whose satisfaction or violation goes into determining the overall probability value the visual system assigns to possible scene interpretations.

2. In places, though, Shepard seems to downplay the need to demonstrate a significant role for the kinematic constraint in more standard conditions. As he says, "Natural selection has ensured that (under favorable viewing conditions) we generally perceive the transformations that an external object is actually undergoing in the external world, however simple or complex, rigid or nonrigid" (Shepard 1994, p. 7). So perhaps the constraint only determines "the default motions that are internally represented under the unfavorable conditions that provide no information about the motion that actually took place" (1994, p. 7). From this standpoint, worries about the ecological validity of the constraint are not very pressing, but then again the significance of the constraint in explaining ordinary perception would be further diminished.

## 6. Evolution

1. Failure of everyday motion to adhere to a principle of geometry or physics does not rule out the possibility that the visual system is guided or influenced by such an internal constraint. Lack of "ecological validity," nevertheless, does make more puzzling aspects of Shepard's internalization thesis. If the actual movements our ancestors experienced were not by and large instances of the unique path specified by the constraint, what would drive or account for the evolutionary incorporation of the principle? And in what sense are we to understand the constraint as reflecting a worldly regularity?

2. If the kinematic constraint is relatively weak or nonexistent in ordinary situations, an additional issue arises for Shepard's account. For such a lack of influence would suggest that the environment is typically rich enough or sufficient to *force* veridical perception independent of the constraint. This makes it more difficult to explain the biological advantage the kinematic constraint is supposed to convey. On the one hand, the constraint is not needed to perceive most everyday motion that conforms to it, the stimuli are rich enough. On the other hand, the constraint may only hinder perception of actual motions that do not fit its specifications. This last point is especially troublesome, since much of the real motion we do encounter does not traverse a path that is the unique, twisting route prescribed by kinematic geometry.

3. In various places Shepard suggests that psychological explanations that do not take an evolutionary approach are shallow, if not defective. I am not convinced this is so. Although an evolutionary perspective may be provocative and can suggest new problems and new lines of attack, models of visual processing and claims about underlying mechanisms can be formulated and tested quite independently of issues of origin. More to the point, if Shepard's kinematic constraint does play a significant role in perception, it should be of interest, even if an evolutionary internalization account of its development could not be sustained.

ACKNOWLEDGMENTS

Work on this paper was supported by the Zentrum fuer interdisziplinaere Forschung, Bielefeld. I wish to thank the members of the ZiF project, especially M. Atherton, D. Hoffman, L. Maloney, and D. Todorovič for discussing these issues with me.

NOTES

**1.** In Shiffrar and Shepard (1991) comparison judgments of paths of real (i.e., computer simulated) movement are offered as support. I do not think the evidence presented there much affects the issues I raise.

**2.** The problem is raised in Shepard (1984) and mentioned but not pursued in Carlton and Shepard (1990a) and Shepard (1994).



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 629–640
*Printed in the United States of America*

# Generalization, similarity, and Bayesian inference

**Joshua B. Tenenbaum and Thomas L. Griffiths**
*Department of Psychology, Stanford University, Stanford, CA 94305-2130*
**jbt@psych.stanford.edu gruffydd@psych.stanford.edu**
**http://www-psych.stanford.edu/~jbt**
**http://www-psych.stanford.edu/~gruffydd/**

**Abstract:** Shepard has argued that a universal law should govern generalization across different domains of perception and cognition, as well as across organisms from different species or even different planets. Starting with some basic assumptions about natural kinds, he derived an exponential decay function as the form of the universal generalization gradient, which accords strikingly well with a wide range of empirical data. However, his original formulation applied only to the ideal case of generalization from a single encountered stimulus to a single novel stimulus, and for stimuli that can be represented as points in a continuous metric psychological space. Here we recast Shepard's theory in a more general Bayesian framework and show how this naturally extends his approach to the more realistic situation of generalizing from multiple consequential stimuli with arbitrary representational structure. Our framework also subsumes a version of Tversky's set-theoretic model of similarity, which is conventionally thought of as the primary alternative to Shepard's continuous metric space model of similarity and generalization. This unification allows us not only to draw deep parallels between the set-theoretic and spatial approaches, but also to significantly advance the explanatory power of set-theoretic models.

**Keywords:** additive clustering; Bayesian inference; categorization; concept learning; contrast model; features; generalization; psychological space; similarity

## 1. Introduction

Consider the hypothetical case of a doctor trying to determine how a particular hormone, naturally produced by the human body, affects the health of patients. It seems likely that patients with too little of the hormone in their blood suffer negative effects, but so do patients with too much of the hormone. Assume that the possible concentration levels of this hormone can be represented as real numbers between 0 and 100 on some arbitrary measuring scale, and that one healthy patient has been examined and found to have a hormone level of 60. What other hormone levels should the doctor consider healthy?

Now imagine a baby robin whose mother has just given it its first worm to eat. The worms in this robin's environment vary in level of skin pigmentation, and only worms with some intermediate density of pigmentation are good to eat; too dark or too light worms are unhealthy. Finally, suppose for simplicity that robins are capable of detecting shades of worm coloration between 0 and 100 on some arbitrary scale, and that the first worm our baby robin has been given scores a skin pigmentation level of 60. Assuming the mother has chosen a worm that is good to eat, what other pigmentation levels should our baby robin consider good to eat?

These two scenarios are both cases of Shepard's (1987b; 1994) ideal generalization problem: given an encounter with a single stimulus (a patient, a worm) that can be represented as a point in some psychological space (a hormone level or pigmentation level of 60), and that has been found to have some particular consequence (healthy, good to eat), what other stimuli in that space should be expected to have the same consequence? Shepard observes that across a wide variety of experimental situations, including both human and animal subjects, generalization gradients tend to fall off approximately exponentially with distance in an appropriately scaled psychological space (as obtained by multidimensional scaling, or MDS). He then gives a rational probabilistic argument for the origin of this universal law, starting with some basic assumptions about the geometry of natural kinds in psychological spaces, which could be expected to apply equally well to doctors or robins, or even aliens from another galaxy. The argument makes no distinction in principle between conscious, deliberate, "cognitive" inferences, such as the healthy hormone levels scenario, and unconscious, automatic, or "perceptual" inferences, such as the good-to-eat worms scenario, as long as they satisfy the conditions of the ideal generalization problem.

In the opening sentences of his first paper on the uni-

JOSHUA B. TENENBAUM is Assistant Professor of Psychology at Stanford University. In 1999, he received a Ph.D. in Brain and Cognitive Sciences from MIT. His research focuses on learning and inference in humans and machines, with specific interests in concept learning and generalization, similarity, reasoning, causal induction, and learning perceptual representations.
THOMAS L. GRIFFITHS is a doctoral student in the Department of Psychology at Stanford University. His research interests concern the application of mathematical and statistical models to human cognition.

© 2001 Cambridge University Press 0140-525X/01 $12.50

https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

versal law of generalization, Shepard (1987b) invokes Newton's universal law of gravitation as the standard to which he aspires in theoretical scope and significance. The analogy holds more strongly than might have been anticipated. Newton's law of gravitation was expressed in terms of the attraction between two point masses: every object in the universe attracts every other object with a force directed along the line connecting their centers of mass, proportional to the product of their masses and inversely proportional to the square of their separation. However, most of the interesting gravitational problems encountered in the universe do not involve two point masses. In order to model real-world gravitational phenomena, physicists following Newton have developed a rich theory of classical mechanics that extends his law of gravitation to address the interactions of multiple, arbitrarily extended bodies.

Likewise, Shepard formulated his universal law with respect to generalization from a single encountered stimulus to a single novel stimulus, and he assumed that stimuli could be represented as points in a continuous metric psychological space. However, many of the interesting problems of generalization in psychological science do not fit this mold. They involve inferences from multiple examples, or stimuli that are not easily represented in strictly spatial terms. For example, what if our doctor observes the hormone levels of not one but three healthy patients: 60, 30, and 50. How should that change the generalization gradient? Or what if the same numbers had been observed in a different context, as examples of a certain mathematical concept presented by a teacher to a student? Certain features of the numbers that were not salient in the hormone context, such as being even or being multiples of ten, now become very important in a mathematical context. Consequently, a simple one-dimensional metric space representation may no longer be appropriate: 80 may be more likely than 47 to be an instance of the mathematical concept exemplified by 60, 30, and 50, while given the same examples in the hormone context, 47 may be more likely than 80 to be a healthy level. Just as physicists now see Newton's original two-point-mass formulation as a special case of the more general classical theory of gravitation, so would we like a more general theory of generalization, which reduces to Shepard's original two-points-in-psychological-space formulation in the appropriate special cases, but which extends his approach to handle generalization from multiple, arbitrarily structured examples.

In this article we outline the foundations of such a theory, working with the tools of Bayesian inference and in the spirit of rational analysis (Anderson 1990; Chater & Oaksford 1998; 1999; Marr 1982). Much of our proposal for extending Shepard's theory to the cases of multiple examples and arbitrary stimulus structures has already been introduced in other papers (Griffiths & Tenenbaum 2000; Tenenbaum 1997; 1999a; 1999b; Tenenbaum & Xu 2000). Our goal here is to make explicit the link to Shepard's work and to use our framework to make connections between his work and other models of learning (Feldman 1997; Gluck & Shanks 1994; Haussler et al. 1994; Kruschke 1992; Mitchell 1997), generalization (Heit 1998; Nosofsky 1986), and similarity (Chater & Hahn 1997; Medin et al. 1993; Tversky 1997). In particular, we will have a lot to say about how our generalization of Shepard's theory relates to Tversky's (1977) well-known set-theoretic models of similarity. Tversky's set-theoretic approach and Shepard's metric space approach are often considered the two classic – and classically opposed – theories of similarity and generalization. By demonstrating close parallels between Tversky's approach and our Bayesian generalization of Shepard's approach, we hope to go some way towards unifying these two theoretical approaches and advancing the explanatory power of each.

The plan of our article is as follows. In section 2, we recast Shepard's analysis of generalization in a more general Bayesian framework, preserving the basic principles of his approach in a form that allows us to apply the theory to situations with multiple examples and arbitrary (nonspatially represented) stimulus structures. Sections 3 and 4 describe those extensions, and section 5 concludes by discussing some implications of our theory for the internalization of perceptual-cognitive universals.

## 2. A Bayesian framework for generalization

Shepard (1987b) formulates the problem of generalization as follows. We are given one example, $x$, of some consequence $C$, such as a "healthy person" or a "good-to-eat worm." We assume that $x$ can be represented as a point in a continuous metric psychological space, such as the one-dimensional space of hormone levels between 0 and 100, and that $C$ corresponds to some region – the *consequential region* – of that space. Our task is then to infer the probability that some newly encountered object $y$ will also be an instance of $C$, that is, that $y$ will fall in the consequential region for $C$. Formalizing this induction problem in probabilistic terms, we are asking for $p(y \in C|x)$, the conditional probability that $y$ falls under $C$ given the observation of the example $x$.

The theory of generalization that Shepard develops and that we will extend here can best be understood by considering how it addresses three crucial questions of learning (after Chomsky 1986):

1. What constitutes the learner's knowledge about the consequential region?
2. How does the learner use that knowledge to decide how to generalize?
3. How can the learner acquire that knowledge from the example encountered?

Our commitment to work within the paradigm of Bayesian probabilistic inference leads directly to rational answers for each of these questions. The rest of this section presents these answers and illustrates them concretely using the hormone or pigmentation levels tasks introduced above. Our main advance over Shepard's original analysis comes in introducing the *size principle* (Tenenbaum 1997; 1999a; 1999b) for scoring hypotheses about the true consequential region based on their size, or specificity. Although it makes little difference for the simplest case of generalization studied by Shepard, the size principle will provide the major explanatory force when we turn to the more realistic cases of generalizing from multiple examples (sect. 3) with arbitrary structure (sect. 4).

### 2.1. What constitutes the learner's knowledge about the consequential region?

The learner's knowledge about the consequential region is represented as a probability distribution $p(h|x)$ over an a priori-specified *hypothesis space* $\mathcal{H}$ of possible consequen-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tial regions $h \in \mathcal{H}$. $\mathcal{H}$ forms a set of exhaustive and mutually exclusive possibilities; that is, one and only one element of $\mathcal{H}$ is assumed to be the true consequential region for $C$ (although the different candidate regions represented in $\mathcal{H}$ may overlap arbitrarily in the stimuli that they include). The learner's background knowledge, which may include both domain-specific and domain-general components, will often translate into constraints on which subsets of objects belong to $\mathcal{H}$. Shepard (1994) suggests the general constraint that consequential regions for basic natural kinds should correspond to connected subsets of psychological space. Applying the connectedness constraint to the domains of hormone levels or worm pigmentation levels, where the relevant stimulus spaces are one-dimensional continua, the hypothesis spaces would consist of intervals, or ranges of stimuli between some minimum and maximum consequential levels. Figure 1 shows a number of such intervals which are consistent with the single example of 60. For simplicity, we have assumed in Figure 1 that only integer stimulus values are possible, but in many cases both the stimulus and hypothesis spaces will form true continua.

At all times, the learner's knowledge about the consequential region consists of a probability distribution over $\mathcal{H}$. Prior to observing $x$, this distribution is the prior probability $p(h)$; after observing $x$, it is the posterior probability $p(h|x)$. As probabilities, $p(h)$ and $p(h|x)$ are numbers between 0 and 1 reflecting the learner's degree of belief that $h$ is in fact the true consequential region corresponding to $C$. In Figure 1, $p(h|x)$ for each $h$ is indicated by the thickness (height) of the corresponding bar. The probability of any $h$ that does not contain $x$ will be zero, because it cannot be the true consequential region if it does not contain the one observed example. Hence, Figure 1 shows only hypotheses consistent with $x = 60$.

### 2.2. How does the learner use that knowledge to decide how to generalize?

The generalization function $p(y \in C|x)$ is computed by summing the probabilities $p(h|x)$ of all hypothesized consequential regions that contain $y$:[1]

$$p(y \in C|x) = \sum_{h:y \in h} p(h|x). \quad (1)$$

We refer to this computation as *hypothesis averaging*, because it can be thought of as averaging the predictions that each hypothesis makes about $y$'s membership in $C$, weighted by the posterior probability of that hypothesis. Because $p(h|x)$ is a probability distribution, normalized to sum to 1 over all $h \in \mathcal{H}$, the structure of Equation 1 ensures that $p(y \in C|x)$ will always lie between 0 and 1. In general, the hypothesis space need not be finite or even countable. In the case of a continuum of hypotheses, such as the space of all intervals of real numbers, all probability distributions over $\mathcal{H}$ become probability densities and the sums over $\mathcal{H}$ (in Equations 1 and following) become integrals.

The top panel of Figure 1 shows the generalization gradient that results from averaging the predictions of the integer-valued hypotheses shown below, weighted by their

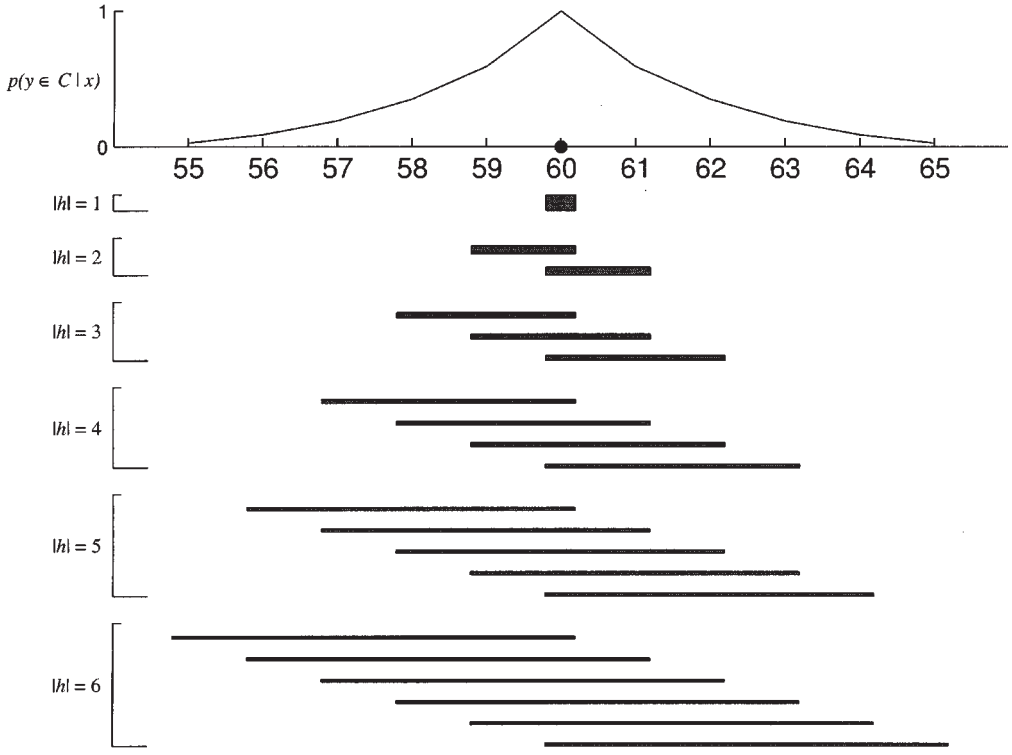



Figure 1. An illustration of the Bayesian approach to generalization from $x = 60$ in a one-dimensional psychological space (inspired by Shepard 1989, August). For the sake of simplicity, only intervals with integer-valued endpoints are shown. All hypotheses of a given size are grouped together in one bracket. The thickness (height) of the bar illustrating each hypothesis $h$ represents $p(h|x)$, the learner's degree of belief that $h$ is the true consequential region given the observation of $x$. The curve at the top of the figure illustrates the gradient of generalization obtained by integrating over just these consequential regions. The profile of generalization is always concave regardless of what values $p(h|x)$ takes on, as long as all hypotheses of the same size (in one bracket) take on the same probability.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

probabilities. Note that the probability of generalization equals 1 only for $y = x$, when every hypothesis containing $x$ also contains $y$. As $y$ moves further away from $x$, the number of hypotheses containing $x$ that also contain $y$ decreases, and the probability of generalization correspondingly decreases. Moreover, Figure 1 shows the characteristic profile of Shepard's "universal" generalization function: concave, or negatively accelerated as $y$ moves away from $x$. If we were to replace the integer-valued interval hypotheses with the full continuum of all real-valued intervals, the sum in Equation 1 would become an integral, and the piecewise linear gradient shown in Figure 1 would become a smooth function with a similar concave profile, much like those depicted in the top panels of Figures 2 and 3.

Figure 1 demonstrates that Shepard's approximately exponential generalization gradient emerges under one particular assignment of $p(h|x)$, but it is reasonable to ask how sensitive this result is to the choice of $p(h|x)$. Shepard (1987b) showed that the shape of the gradient is remarkably insensitive to the probabilities assumed. As long as the probability distribution $p(h|x)$ is *isotropic,* that is, independent of the location of $h$, the generalization function will always have a concave profile. The condition of isotropy is equivalent to saying that $p(h|x)$ depends only on $|h|$, the size of the region $h$; notice how this constraint is satisfied in Figure 1.

### 2.3. How can the learner acquire that knowledge from the example encountered?

After observing $x$ as an example of the consequence $C$, the learner updates her beliefs about the consequential region from the prior $p(h)$ to the posterior $p(h|x)$. Here we consider how a rational learner arrives at $p(h|x)$ from $p(h)$, through the use of Bayes' rule. We will not have much to say about the origins of $p(h)$ until section 5; Shepard (1987b) and Tenenbaum (1999a; 1999b) discuss several reasonable alternatives for the present scenarios, all of which are isotropic and assume little or no knowledge about the true consequential region.

Bayes' rule couples the posterior to the prior via the *likelihood,* $p(x|h)$, the probability of observing the example $x$ given that $h$ is the true consequential region, as follows:

$$p(h|x) = \frac{p(x|h)p(h)}{p(x)} \tag{2}$$

$$= \frac{p(x|h)p(h)}{\sum_{h' \in \mathcal{H}} p(x|h')p(h')}. \tag{3}$$

What likelihood function we use is determined by how we think the process that generated the example $x$ relates to the true consequential region for $C$. Shepard (1987b) argues for a default assumption that the example $x$ and consequential region $C$ are sampled independently, and $x$ just happens to land inside $C$. This assumption is standard in the machine learning literature (Haussler et al. 1994; Mitchell 1997), and also maps onto Heit's (1998) recent Bayesian analysis of inductive reasoning. Tenenbaum (1997; 1999a) argues that under many conditions, it is more natural to treat $x$ as a random positive example of $C$, which involves the stronger assumption that $x$ was explicitly sampled from $C$. We refer to these two models as weak sampling and strong sampling, respectively.

Under weak sampling, the likelihood just measures in a binary fashion whether or not the hypothesis is consistent with the observed example:

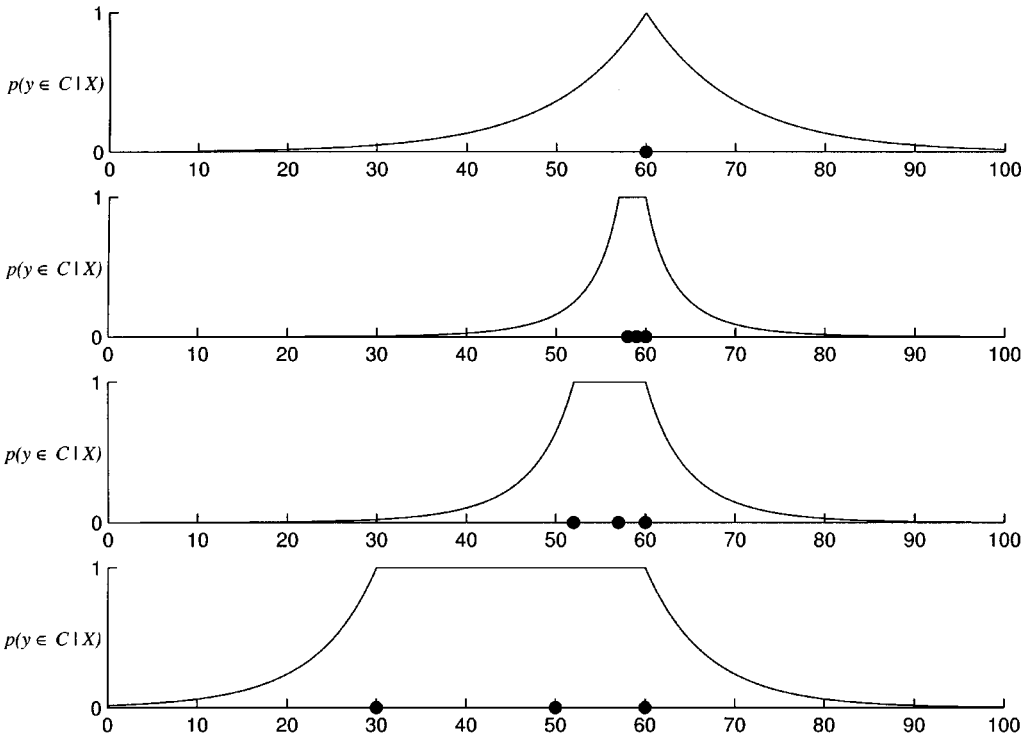



Figure 2. The effect of example variability on Bayesian generalization (under the assumptions of strong sampling and an Erlang prior, $\mu = 10$). Filled circles indicate examples. The first curve is the gradient of generalization with a single example, for the purpose of comparison. The remaining graphs show that the range of generalization increases as a function of the range of examples.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

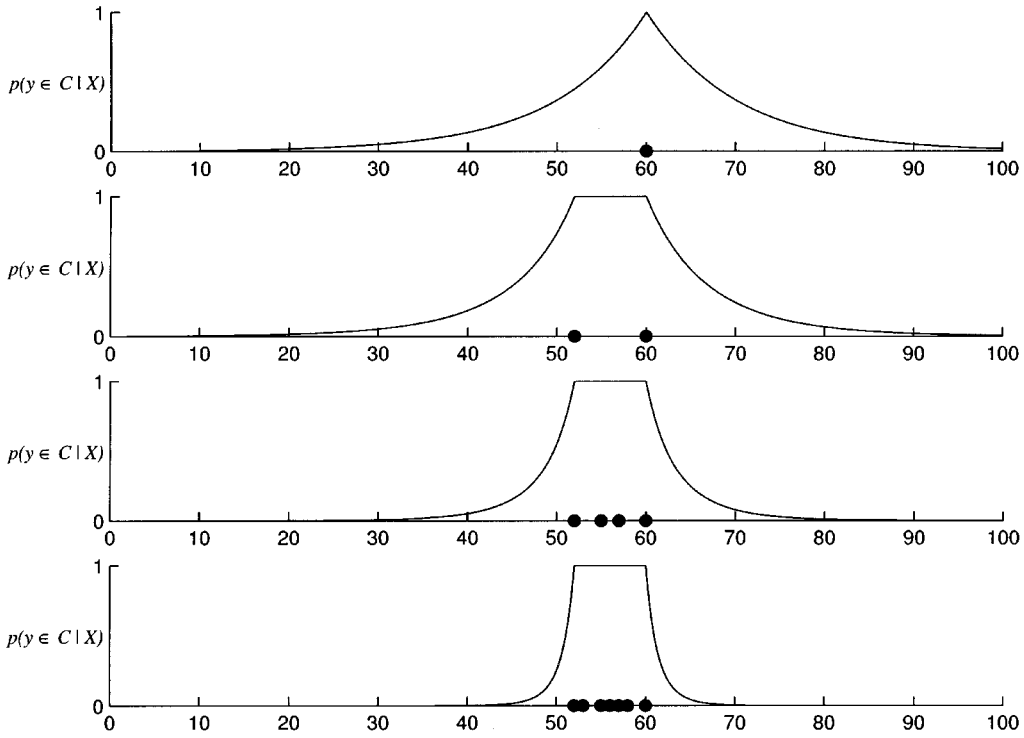



Figure 3. The effect of the number of examples on Bayesian generalization (under the assumptions of strong sampling and an Erlang prior, $\mu = 10$). Filled circles indicate examples. The first curve is the gradient of generalization with a single example, for the purpose of comparison. The remaining graphs show that the range of generalization decreases as a function of the number of examples.

$$p(x|h) = \begin{cases} 1 & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases} \quad \text{[weak sampling]}. \qquad (4)$$

Under strong sampling, the likelihood is more informative. Assuming $x$ is sampled from a uniform distribution over the objects in $h$, we have:

$$p(x|h) = \begin{cases} \frac{1}{|h|} & \text{if } x \in h \\ 0 & \text{otherwise} \end{cases} \quad \text{[strong sampling]}, \qquad (5)$$

where $|h|$ indicates the size of the region $h$. For discrete stimulus spaces, $|h|$ is simply the cardinality of the subset corresponding to $h$. For continuous spaces such as the hormone or pigmentation levels, the likelihood becomes a probability density and $|h|$ is the *measure* of the hypothesis – in one dimension, just the length of the interval.[2] Equation 5 implies that smaller, more specific hypotheses will tend to receive higher probabilities than larger, more general hypotheses, even when both are equally consistent with the observed consequential stimulus. We will call this tendency the *size principle*. It is closely related to principles of *genericity* that have been proposed in models of visual perception and categorization (Feldman 1997; Knill & Richards 1996). Figure 1 depicts the application of the size principle graphically.

Note that both Equations 4 and 5 are isotropic, and thus the choice between strong sampling and weak sampling has no effect on Shepard's main result that generalization gradients are universally concave. However, as we now turn to look at the phenomena of generalization from multiple stimuli with arbitrary, nonspatially represented structures, we will see that the size principle implied by strong sampling carries a great deal of explanatory power not present in Shepard's original analysis.

## 3. Multiple examples

In this section, we extend the above Bayesian analysis to situations with multiple consequential examples. Such situations arise quite naturally in the generalization scenarios we have already discussed. For instance, how should our doctor generalize after observing hormone levels of 60, 30, and 50 in three healthy patients? We first discuss some basic phenomena that arise with multiple examples and then turn to the extension of the theory. Finally, we compare our approach to some alternative ways in which Shepard's theory has been adapted to apply to multiple examples.

### 3.1. Phenomena of generalization from multiple examples

We focus on two classes of phenomena: the effects of example variability and the number of examples.

**3.1.1. Example variability.** All other things being equal, the lower the variability in the set of observed examples, the lower the probability of generalization outside their range. The probability that 70 is a healthy hormone level seems greater given the three examples {60, 50, 30} than given the three examples {60, 57, 52}, and greater given {60, 57, 52} than given {60, 58, 59}. Effects of exemplar variability on generalization have been documented in several other categorization and inductive inference tasks (Fried & Holyoak 1984; Osherson et al. 1990; Rips 1989).

**3.1.2. Number of examples.** All other things being equal, the more examples observed within a given range, the lower the probability of generalization outside that range. The



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

probability that 70 is a healthy hormone level seems greater given the two examples {60, 52} than given the four examples {60, 52, 57, 55}, and greater given {60, 52, 57, 55} than given {60, 52, 57, 55, 58, 55, 53, 56}. This effect is most dramatic when there is very little variability in the observed examples. Consider the three sets of examples {60}, {60, 62, 61}, and {60, 62, 61, 62, 60, 62, 60, 61}. With just two more examples, the probability of generalizing to 70 from {60, 62, 61} already seems much lower than given {60} alone, and the probability given {60, 62, 61, 62, 60, 62, 60, 61} seems close to zero.

### 3.2. Extending the theory

Let $X = \{x_1, \ldots x_n\}$ denote a sequence of $n$ examples of some consequence $C$, and let $y$ denote a novel object for which we want to compute the probability of generalizing, $p(y \in C|X)$. All we have to do to make the theory of section 2 applicable here is to replace "$x$," wherever it appears, with "$X$," and to adopt the assumption of strong sampling rather than Shepard's original proposal of weak sampling. The rest of the formalism is unchanged. The only complication this introduces comes in computing the likelihood $p(X|h)$. If we make the simplifying assumption that the examples are sampled independently of each other (a standard assumption in Bayesian analysis), then Equation 5 becomes:

$$p(X|h) = \prod_i p(x_i|h) \tag{6}$$

$$= \begin{cases} \frac{1}{|h|^n} & \text{if } x_1, \ldots, x_n \in h \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Hence the size principle of Equation 5 has been generalized to include the influence of *n:* smaller hypotheses receive higher likelihoods than larger hypotheses, by a factor that increases exponentially with the number of examples observed. Figures 2 and 3 depict the Bayesian gradients of generalization that result for several different numbers and ranges of examples, assuming $p(X|h)$ based on strong sampling and an Erlang distribution (Shepard 1987b) for $p(h)$. In addition to showing the universal concave profile, these gradients display the appropriate sensitivity to the number and variability of examples.

To understand how the size principle generates these effects, consider how Equation 7 weights two representative hypotheses: $h_0$, the smallest interval containing all the examples in $X$, and $h_1$, a broader interval centered on $h_0$ but extending by $d/2$ units on either side, so that $|h_1| = |h_0| + d$. After observing $n$ examples, the relative probabilities are proportional to the likelihood ratio:

$$\mathcal{L} = \frac{p(X|h_1)}{p(X|h_0)} = \left[\frac{1}{1 + d/|h_0|}\right]^n. \tag{8}$$

$\mathcal{L}$ is always less than 1, because $d$ and $|h_0|$ are both positive. As $|h_0|$ increases, but the other quantities remain fixed, $\mathcal{L}$ increases. Thus, as we see in Figure 2, the relative probability that $C$ extends a given distance $d$ beyond the examples increases as the range spanned by the examples increases. As $n$ increases while the other quantities remain fixed, $\mathcal{L}$ quickly approaches 0. Thus, as we see in Figure 3, the probability that $C$ extends a distance $d$ beyond the examples rapidly decreases as the number of examples increases within a fixed range. The tighter the examples, the smaller $|h_0|$ is, and the faster $\mathcal{L}$ decreases with increasing $n$, thus accounting for the interaction between these two factors pointed to earlier.

We can also now see why Shepard's original assumption of weak sampling would not generate these phenomena. Under weak sampling, the likelihoods of any two consistent hypotheses are always both 1. Thus $\mathcal{L} = 1$ always, and neither the range nor the number of examples have any effect on how hypotheses are weighted. In general, we expect that both strong sampling and weak sampling models will have their uses. Real-world learning situations may often require a combination of the two, if some examples are generated by mere observation of consequential stimuli (strong sampling) and others by trial-and-error exploration (weak sampling).

Figure 4 illustrates an extension to generalizing in two separable dimensions, such as inferring the healthy levels of two independent hormones (for more details, see Tenenbaum 1999b). Following Shepard (1987b), we assume that the consequential regions correspond to axis-aligned rectangles in this two-dimensional space, with independent priors in each dimension. Then, as shown in Figure 4, the size principle acts to favor generalization along those dimensions for which the examples have high variability and to restrict generalization along dimensions for which they have low variability. Tenenbaum (1999b) reports data from human subjects that are consistent with these predictions for a task of estimating the healthy levels of two biochemical compounds. More studies need to be done to test these predictions in multidimensional perceptual spaces of the sort with which Shepard has been most concerned.

### 3.3. Alternative approaches

A number of other computational models may be seen as alternative methods of extending Shepard's approach to the case of multiple examples, but only the framework we describe here preserves what we take to be the two central features of Shepard's original analysis: a hypothesis space of possible consequential regions and a Bayesian inference procedure for updating beliefs about the true consequential region. The standard exemplar models of classification (e.g., Nosofsky 1986; 1998a) take Shepard's exponential law of generalization as a primitive, used to justify the assumption that exemplar activation functions decay exponentially with distance in psychological space. A different approach is based on connectionist networks (Gluck 1991; Shanks & Gluck 1994; Shepard & Kannapan 1990; Shepard & Tenenbaum 1991), in which input or hidden units represent consequential regions, and error-driven learning – rather than Bayesian inference – is used to adjust the weights from consequential region inputs to response outputs. A third class of models (Kruschke 1992; Love & Medin 1998) combines aspects of the first two, by embedding Shepard's exponential law within the activation functions of hidden units in a connectionist network for classification learning.

Space does not permit a full comparison of the various alternative models with our proposals. One important point of difference is that for most of these models, the generalization gradients produced by multiple examples of a given consequence are essentially just superpositions of the exponential decay gradients produced by each individual example. Consequently, those models cannot easily explain


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

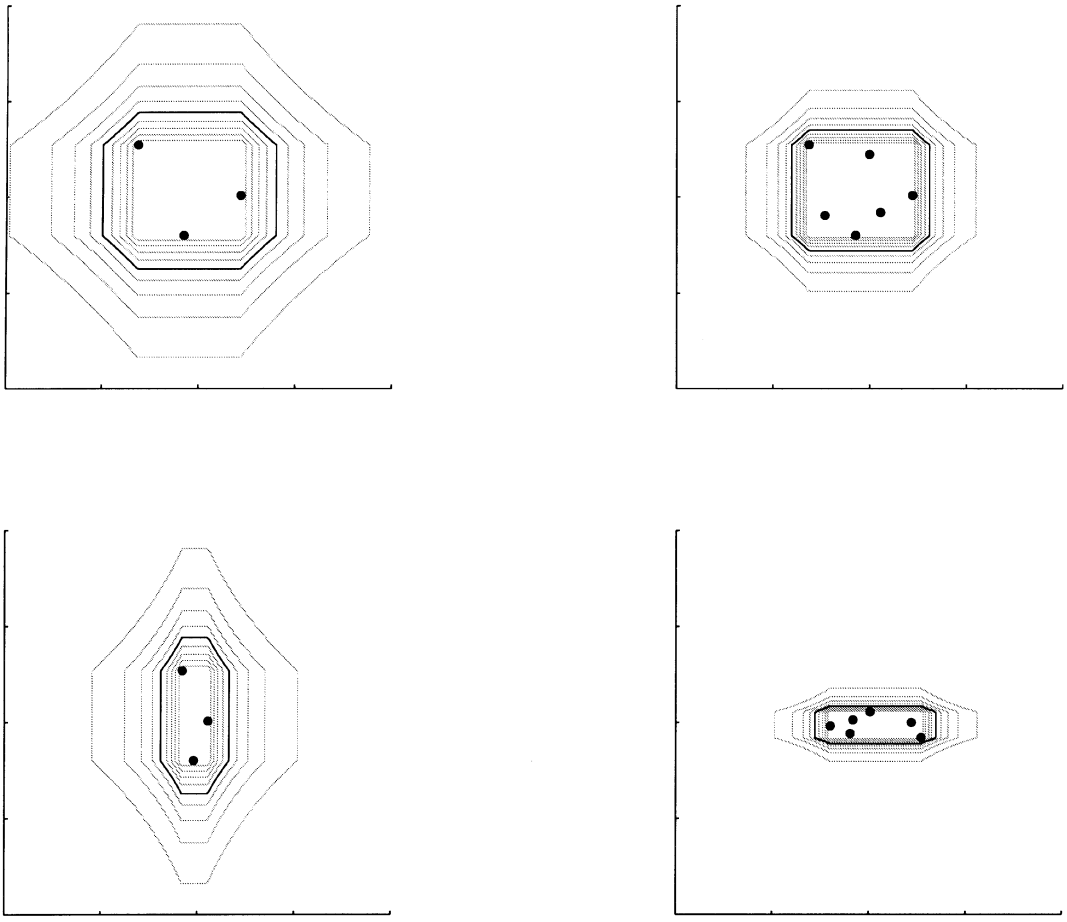


Figure 4. Bayesian generalization from multiple examples in two separable dimensions. Examples are indicated by filled circles. Contours show posterior probability, in increments of 0.1. Black contours illustrate the points at which $p(y \in C|X) = 0.5$. The range of generalization is affected by both the number of examples and the variability along a given dimension.

the phenomena discussed above, in which encountering additional consequential stimuli causes the probability of generalizing to some new stimulus to *decrease,* even when the additional examples are more similar to the new stimulus than the original example was. Exemplar and exemplar/connectionist hybrid models are frequently equipped with variable "attentional weights" that scale distances along a given input dimension by a greater or lesser amount, in order to produce variations in the contours of generalization like those in Figure 4. Such models could account for our phenomena by postulating that a dimension's length scale is initially large and decreases as the number of examples increases or the variability of the examples decreases, but nothing in the formal structure of these models necessarily implies such a mechanism. Our Bayesian analysis, in contrast, necessarily predicts these effects as rational consequences of the size principle.

## 4. Arbitrary stimulus structure

Shepard (1987b) assumed that objects can be represented as points in a continuous metric psychological space, and that the consequential subsets correspond to regions in that space with some convenient properties, such as connectedness or central symmetry. In general, though, we do not need to assume that the hypothesized consequential subsets correspond to regions in any continuous metric space; the notion of a consequential subset is sufficient for defining a Bayesian account of generalization. In this section we examine how arbitrary, nonspatially represented stimulus structures are modeled within the Bayesian framework.

Several authors, including Shepard himself, have described extensions of the original theory of generalization to conjunctive feature structures, in which objects are represented in terms of the presence or absence of primitive binary features and the possible consequential subsets consist of all objects sharing different conjunctions of features. For these cases, generalization gradients can still be shown to follow an exponential-like decay function of some appropriately defined distance measure (Gluck 1991; Russell 1988; Shepard 1989; 1994). However, the Bayesian analysis of generalization is more widely applicable than this. As we will show here, the analysis applies even when there is no independent notion of distance between stimuli and nothing like an exponential gradient emerges from the sum over consequential regions.

To motivate our analysis, consider a new generalization scenario. A computer has been programmed with a variety of simple mathematical concepts defined over the integers 1–100 – subsets of numbers that share a common, mathematically consequential property such as "even number,"



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

"power of two," or "square number." The computer will select one of these subsets at random, choose one or more numbers at random from that subset to show you as examples, and then quiz you by asking if certain other numbers belong to this same concept. Suppose that the number 60 is offered as one example of a concept the computer has chosen. What is the probability that the computer will accept 50? How about 51, 47, or 80? Syntactically, this task is almost identical to the hormone levels scenario above. But now, instead of generalization following a monotonic function of proximity in numerical magnitude, it seems more likely to follow some measure of mathematical similarity. For instance, the number 60 shares more mathematical properties with 50 than with 51, making 50 perhaps a better bet than 51 to be accepted given the one example of 60, even though 51 is closer in magnitude to 60 and therefore a better bet for the doctor trying to determine healthy hormone levels.

In our Bayesian framework, the difference between the two scenarios stems from the very different consequential subsets (elements of $\mathcal{H}$) that are considered. For the doctor, knowing something about healthy levels of hormones in general, it is quite natural to assume that the true consequential subset corresponds to some unknown interval, which gives rise to a generalization function monotonically related to proximity in magnitude. To model the number game, we can identify each mathematical property that the learner knows about with a possible consequential subset in $\mathcal{H}$. Figure 5 shows a generalization function that results under a set of 33 simple hypotheses, as calculated from the size principle (Eq. 5) and hypothesis averaging (Eq. 1). The generalization function appears much more jagged than in Figures 1–3 because the mathematical hypothesis space does not respect proximity in the dimension of numerical magnitude (corresponding to the abscissa of the figures). More generally, numerical cognition may incorporate both the spatial, magnitude properties as well as the nonspatial, mathematical properties of numbers. To investigate the nature of mental representations of numbers, Shepard et al. (1975) collected human similarity judgments for all pairs of integers between 0 and 9, under a range of different contexts. By submitting these data to an additive clustering analysis (Shepard & Arabie 1979; Tenenbaum 1996), we can construct the hypothesis space of consequential subsets that best accounts for people's similarity judgments. Table 1 shows that two kinds of subsets occur in the best-fitting additive clustering solution (Tenenbaum 1996): numbers sharing a common mathematical property, such as {2, 4, 8} and {3, 6, 9}, and consecutive numbers of similar magnitude, such as {1, 2, 3, 4} and {2, 3, 4, 5, 6}. Tenenbaum (2000) studied how people generalized concepts in a version of the number game that made both mathematical and magnitude properties salient. He found that a Bayesian model using a hypothesis space inspired by these additive clustering results, but defined over the integers 1–100, yielded an excellent fit to people's generalization judgments. The same flexibility in hypothesis space structure that allows the Bayesian framework to model both the spatial hormone level scenario and the nonspatial number game scenario there allows it to model generalization in a more generic context, by hypothesizing a mixture of consequential subsets for both spatial, magnitude properties and nonspatial, mathematical properties. In fact, we can define a Bayesian generalization function not just for spatial, featural, or simple hybrids of these representations, but for almost any collection of hypothesis subsets $\mathcal{H}$ whatsoever. The only restriction is that we be able to define a prior probability measure (discrete or continuous) over $\mathcal{H}$, and a measure over the space of objects, required for strong sampling to make sense. Even without a measure over the space of objects, a Bayesian analysis using weak sampling will still be possible.

### 4.1. Relations between generalization and set-theoretic models of similarity

Classically, mathematical models of similarity and generalization fall between two poles: continuous metric space models such as in Shepard's theory, and set-theoretic matching models such as Tversky's (1977) contrast model. The latter strictly include the former as a special case, but are most commonly applied in domains where a set of discrete conceptual features, as opposed to a low-dimensional continuous space, seems to provide the most natural stimulus representation (Shepard 1980). Our number game is such a domain, and indeed, when we generalize Shepard's Bayesian analysis from consequential regions in continuous metric spaces to apply to arbitrary consequential subsets, the model comes to look very much like a version of Tversky's set-theoretic models. Making this connection explicit allows us not only to unify the two classically opposing approaches to similarity and generalization, but also to explain

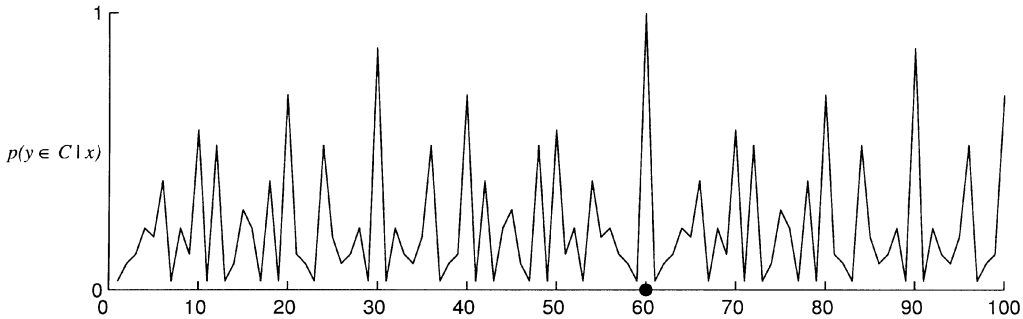



Figure 5. Bayesian generalization in the number game, given one example $x = 60$. The hypothesis space includes 33 mathematically consequential subsets (with equal prior probabilities): even numbers, odd numbers, primes, perfect squares, perfect cubes, multiples of a small number (3–10), powers of a small number (2–10), numbers ending in the same digit (1–9), numbers with both digits equal, and all numbers less than 100.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

some significant aspects of similarity that Tversky's original treatment did not attempt to explain.

Tversky's (1977) contrast model expresses the similarity of $y$ to $x$ as

$$S(y,x) = \theta f(\mathcal{Y} \cap \mathcal{X}) - \alpha f(\mathcal{Y} - \mathcal{X}) - \beta f(\mathcal{X} - \mathcal{Y}), \quad (9)$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the feature sets representing $x$ and $y$, respectively, $f$ denotes some measure over the feature sets, and $\theta, \alpha, \beta$ are free parameters of the model. Similarity thus involves a contrast between the *common* features of $y$ and $x$, $\mathcal{Y} \cap \mathcal{X}$, and their *distinctive* features, those possessed by $y$ but not $x$, $\mathcal{Y} - \mathcal{X}$, and those possessed by $x$ but not $y$, $\mathcal{X} - \mathcal{Y}$. Tversky also suggested an alternative form for the matching function, the *ratio* model, which can be written as

$$S(y,x) = 1 \Big/ \left[1 + \frac{\alpha f(\mathcal{Y} - \mathcal{X}) + \beta f(\mathcal{X} - \mathcal{Y})}{f(\mathcal{Y} \cap \mathcal{X})}\right]. \quad (10)$$

The ratio model is remarkably similar to our Bayesian model of generalization, which becomes particularly apparent when the Bayesian model is expressed in the following form (mathematically equivalent to Eq. 1):

$$p(y \in C|x) = 1 \Big/ \left[1 + \frac{\sum_{h:x \in h, y \notin h} p(h,x)}{\sum_{h:x,y \in h} p(h,x)}\right]. \quad (11)$$

Here, $p(h, x) = p(x|h)p(h)$ represents the weight assigned to hypothesis $h$ in light of the example $x$, which depends on both the prior and the likelihood. The bottom sum ranges over all hypotheses that include both $x$ and $y$, while the top sum ranges over only those hypotheses that include $x$ but do not include $y$. If we identify each feature $k$ in Tversky's framework with a hypothesized subset $h$, where an object belongs to $h$ if and only if it possesses feature $k$, and if we make the standard assumption that the measure $f$ is additive, then the Bayesian model as expressed in Equation 11 corresponds formally to the ratio model with $\alpha = 0$, $\beta = 1$. It is also monotonically related to the contrast model, under the same parameter settings.

Interpreting this formal correspondence between our Bayesian model of generalization and Tversky's set-theoretic models of similarity is complicated by the fact that in general the relation between similarity and generalization is not well understood. A number of authors have proposed that similarity is the more primitive cognitive process and forms (part of) the basis for our capacity to generalize inductively (Goldstone 1994; Osherson et al. 1990; Quine 1969; Rips 1975; Smith 1989). But from the standpoint of reverse-engineering the mind and explaining why human similarity or generalization computations take the form that they do, a satisfying theory of similarity is more likely to depend upon a theory of generalization than vice versa. The problem of generalization can be stated objectively and given a principled rational analysis, while the question of how similar two objects are is notoriously slippery and underdetermined (Goodman 1972). We expect that, depending on the context of judgment, the similarity of $y$ to $x$ may involve the probability of generalizing from $x$ to $y$, or from $y$ to $x$, or some combination of those two. It may also depend on other factors altogether. Qualifications aside, interesting consequences nonetheless follow just from the hypothesis that similarity somehow depends on generalization, without specifying the exact nature of the dependence.

**4.1.1. The syntax of similarity.** Most fundamentally, our Bayesian analysis provides a rational basis for the qualitative form of set-theoretic models of similarity. For instance, it explains why similarity should in principle depend on both the common *and* the distinctive features of objects. Tversky (1977) asserted as an axiom that similarity is a function of both common and distinctive features, and he presented some empirical evidence consistent with that assumption, but he did not attempt to explain why it should hold in general. Indeed, there exist both empirical models (Shepard 1980) and theoretical arguments (Chater & Hahn 1997) that have successfully employed only common or distinctive features. Our rational analysis (Eq. 11), in contrast, explains why both kinds of features should matter in general, under the assumption that similarity depends on generalization. The more hypothesized consequential subsets that contain both $x$ and $y$ (common features of $x$ and $y$), relative to the number that contain only $x$ (distinctive features of $x$), the higher the probability that a subset known to contain $x$ will also contain $y$.

Along similar lines, the hypothesis that similarity depends in part on generalization explains why similarity may in principle be an asymmetric relationship, that is, why the similarity of $x$ to $y$ may differ from the similarity of $y$ to $x$. Tversky (1977) presented compelling demonstrations of such asymmetries and showed that they could be modeled in his set-theoretic framework if the two subsets of distinctive features $\mathcal{X} - \mathcal{Y}$ and $\mathcal{Y} - \mathcal{X}$ have different measures under $f$ and are given different weights in Equations 9 or 10. But Tversky's formal theory does not explain *why* those two subsets should be given different weights; it merely allows this as one possibility. In contrast, the probability of generalizing from $x$ to $y$ is intrinsically an asymmetric function, depending upon the distinctive features of $x$ but not those of $y$. Likewise, the probability of generalizing from $y$ to $x$ depends only on the distinctive features of $y$, not those of $x$. To the extent that similarity depends on either or both of these generalization probabilities, it inherits their intrinsic asymmetry. Note that generalization can still be symmetric, when the distinctive features of $x$ and $y$ are equal in number and weight. This condition holds in the spatial scenarios considered above and in Shepard's work, which (not coincidentally) are also the domains in which similarity is found to be most nearly symmetric (Tversky 1977).

Finally, like Shepard's analysis of generalization, Tversky's contrast model was originally defined only for the comparison of two individual objects. However, our Bayesian framework justifies a natural extension to the problem of computing the similarity of an object $y$ to a set of objects $X = \{x_1, \ldots x_n\}$ as a whole, just as it did for Shepard's theory in section 3. Heit (1997a) proposed on intuitive grounds that the contrast model should still apply in this situation, but with the feature set $\mathcal{X}$ for the examples as a whole identified with $\cap_{i=1}^{n} \mathcal{X}_i$, the intersection of the feature sets of all the individual examples. Our Bayesian analysis (replacing $x$ with $X$ in Eq. 11) explains why the intersection, as opposed to some other combination mechanism such as the union, is appropriate. Only those hypotheses consistent with all the examples in $X$ – corresponding to those features belonging to the intersection of all the feature sets $\mathcal{X}_i$ – receive non-zero likelihood under Equation 7.

**4.1.2. The semantics of similarity.** Perhaps the most persistent criticisms of the contrast model and its relatives fo-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

cus on semantic questions: What qualifies as a feature? What determines the feature weights? How do the weights change across judgment contexts? The contrast model has such broad explanatory scope because it allows any kind of features and any feature weights whatsoever, but this same lack of constraint also prevents the model from explaining the origins of the features or weights. Our Bayesian model likewise offers no constraints about what qualifies as a feature, but it does explain some aspects of the origins and the dynamics of feature weights. The Bayesian feature weight $p(h, x) = p(x|h)p(h)$ decomposes into prior and likelihood terms. The prior $p(h)$ is not constrained by our analysis; it can accommodate arbitrary flexibility across contexts but explains none of that flexibility. In contrast, the likelihood $p(x|h)$ is constrained by the assumption of strong sampling to follow the size principle.

One direct implication of this constraint is that, in a given context, features belonging to fewer objects – corresponding to hypotheses with smaller sizes – should be assigned higher weights. This prediction can be tested using additive clustering analyses, which recover a combination of feature extensions and feature weights that best fit a given similarity data set. For instance, the additive clustering analysis of the integers 0–9 presented in Table 1 is consistent with our prediction, with a negative correlation ($r = -0.83$) between the number of stimuli in each cluster and the corresponding feature weights. Similar relationships can be found in several other additive clustering analyses (Arabie & Carroll 1980; Chaturvedi & Carroll 1994; Lee, submitted; Tenenbaum 1996); see Tenenbaum et al. (in preparation) for a comprehensive study. Tversky (1977) proposed several general principles of feature weighting, such as the diagnosticity principle, but he did not explicitly propose a correlation between feature specificity and feature weight, nor was his formal model designed to predict these effects.

A second implication of the size principle is that certain kinds of features should tend to receive higher weights in similarity comparisons, if they systematically belong to fewer objects. Medin et al. (1993) have argued that primitive features are often not as important as are *relational* features, that is, higher-order features defined by relations between primitives. Yet in some cases a relation appears *less* important than a primitive feature. Consider which bottom stimulus, A or B, is more similar to the top stimulus in each panel of Figure 6 (inspired by Medin et al.'s comparisons). In the left panel, the top stimulus shares a primitive feature with B ("triangle on top") and a relational feature with A ("all different shapes"). In an informal survey, 8 out of 10 observers chose B – the primitive feature match – as more similar at first glance. In the right panel, however, a different relation ("all same shape") dominates over the same primitive feature (9 out of 10 different observers chose A as more similar). Goldstone et al. (1989) report several other cases where "same" relations are weighted more highly than "different" relations in similarity comparisons. If similarity depends in part upon Bayesian generalization, then the size principle can explain the relative salience of these features in Figure 6. Let $m$ be the number of distinct shapes (square, triangle, etc.) that can appear in the three positions of each stimulus pattern. Then the consequential subset for "all same shape" contains exactly $m$ distinct stimuli, the subset for "triangle on top" contains $m^2$ stimuli, and the subset for "all different shapes" contains $m(m - 1)(m - 2)$ stimuli. Thus feature saliency is inversely related to subset size, just as we would expect under the size principle. More careful empirical tests of this hypothesis are required, but we conjecture that much of the relative importance of relational features versus primitive features may be explained by their differing specificities.

A final implication arises from the interaction of the size principle with multiple examples. Recall that in generalizing from multiple examples, the likelihood preference for smaller hypotheses increases exponentially in the number of examples (Eq. 7). The same effect can be observed with the weights of features in similarity judgments. For instance, in assessing the similarity of a number to 60, the feature "multiple of ten" may or may not receive slightly greater weight than the feature "even number." But in assessing similarity to the set of numbers {60, 80, 10, 30} as a whole, even though both of those features are equally consistent with the full set of examples, the more specific feature "multiple of ten" appears to be much more salient.

Table 1. *Additive clustering of similarity judgments for the integers 0–9 (from Tenenbaum 1996)*

| Rank | Weight | Stimuli in class | Interpretation |
|---|---|---|---|
| 1 | .444 | 2 4 8 | powers of two |
| 2 | .345 | 0 1 2 | small numbers |
| 3 | .331 | 3 6 9 | multiples of three |
| 4 | .291 | 6 7 8 9 | large numbers |
| 5 | .255 | 2 3 4 5 6 | middle numbers |
| 6 | .216 | 1 3 5 7 9 | odd numbers |
| 7 | .214 | 1 2 3 4 | smallish numbers |
| 8 | .172 | 4 5 6 7 8 | largish numbers |

## 5. Conclusions: Learning, evolution, and the origins of hypothesis spaces

We have described a Bayesian framework for learning and generalization that significantly extends Shepard's theory in two principal ways. In addressing generalization from multiple examples, our analysis is a fairly direct extension of

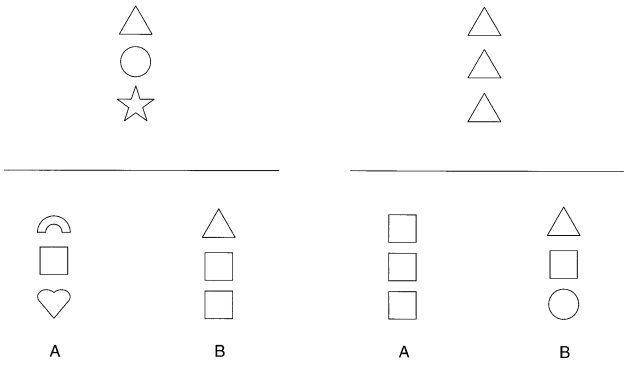



Figure 6. The relative weight of relations and primitive features depends on the size of the set of objects that they identify. Most observers choose B (the primitive feature match) as more similar to the top stimulus in the left panel, but choose A (the relational match) in the right panel, in part because the relation "all same shape" identifies a much smaller subset of objects than the relation "all different shapes."


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Shepard's original ideas, making no substantive additional assumptions other than strong sampling. In contrast, our analysis of generalization with arbitrarily structured stimuli represents a more radical broadening of Shepard's approach, in giving up the notion that generalization is constrained by the metric properties of an evolutionarily internalized psychological space. On the positive side, this step allows us to draw together Tversky's set-theoretic models of similarity and Shepard's continuous metric space models of generalization under a single rational framework, and even to advance the explanatory power of Tversky's set-theoretic models using the same tools – chiefly, the size principle – that we used to advance Shepard's analysis of generalization. Yet it also opens the door to some large unanswered questions, which we close our article by pointing out.

In discussing similarity or generalization with arbitrarily structured stimuli, our Bayesian analysis explains only one piece of the puzzle of how features or hypotheses are weighted. Weights are always a product of both size-based likelihoods and priors, and while the size principle follows rationally from the assumption of strong sampling, the assignment of prior probabilities lies outside the scope of a basic Bayesian analysis. Thus, we can never say anything for certain about the relative weights of any two particular features or hypotheses merely based on their relative sizes; any size difference can always be overruled by a greater difference in prior probability.

The ability of prior probability differences to overrule an opposing size-based likelihood difference is hardly pathological; on the contrary, it is essential in every successful inductive generalization. Consider as a hypothesis in the number game that the computer accepts all multiples of ten, except 20 and 70. "Multiples of ten, except 20 and 70" is slightly more specific than "all multiples of ten," and thus should receive higher probability under the size principle given a set of examples that is consistent with both hypotheses, such as {60, 80, 10, 30}. But obviously, that does not happen in most people's minds. Our Bayesian framework can accommodate this phenomenon by stipulating that while the former hypothesis receives a somewhat higher likelihood, it receives a very much lower prior probability, and thus a significantly lower posterior probability when the prior and likelihood are combined.

It is by now almost a truism that without some reasonable a priori constraints on the hypotheses that learners should consider, there will always be innumerable bizarre hypotheses such as "all multiples of ten, except 20 and 70" that will stand in the way of reasonable inductive generalizations (Goodman 1955; 1972; Mitchell 1997). Trying to determine the nature and origin of these constraints is one of the major goals of much current research (e.g., Medin et al. 1993; Schyns et al. 1998). Shepard's original analysis of generalization was so compelling in part because it proposed answers to these questions: sufficient constraints on the form of generalization are provided merely by the representation of stimuli as points in a continuous metric psychological space (together with the assumption that hypotheses correspond to a suitable family of regions in that space), and our psychological spaces themselves are the products of an evolutionary process that has shaped them optimally to reflect the structure of our environment. In proposing a theory of generalization that allows for arbitrarily structured hypothesis spaces, we owe some account of where those hypothesis spaces and priors might come from. Evolution alone is not sufficient to explain why hypotheses such as "multiples of ten" are considered natural while hypotheses such as "all multiples of ten, except 20 and 70" are not.

The major alternative to evolution as the source of hypothesis space structure is some kind of prior learning. Most directly, prior experience that all and only those objects belonging to some particular subset $h$ tend to possess a number of important consequences may lead learners to increase $p(h)$ for new consequences of the same sort. Unsupervised learning – observation of the properties of objects without any consequential input – may also be extremely useful in forming a hypothesis space for supervised (consequential) learning. Noting that a subset of objects tend to cluster together, to be more similar to each other than to other objects on some primitive features, may increase a learner's prior probability that this subset is likely to share some important but as-yet-unencountered consequence. The machine learning community is now intensely interested in improving the inductive generalizations that a supervised learning agent can draw from a few labeled examples, by building on unsupervised inferences that the agent can draw from a large body of unlabeled examples (e.g., Mitchell 1999; Poggio & Shelton 1999). We expect this to become a critical issue in the near future for cognitive science as well.

Our proposal that the building blocks of Shepard's "perceptual-cognitive universals" come into our heads via learning, and not just evolution, resonates with at least one other contribution to this issue (see Barlow's target article). However, we fundamentally agree with an earlier statement of Shepard's, that "learning is not an alternative to evolution but itself depends on evolution. There can be no learning in the absence of principles of learning; yet such principles, being themselves unlearned, must have been shaped by evolution" (Shepard 1995a, p. 59). Ultimately, we believe that it may be difficult or impossible to separate the contributions that learning and evolution each make to the internalization of world structure, given the crucial role that each process plays in making the other an ecologically viable means of adaptation. Rather, we think that it may be more worthwhile to look for productive synergies of the two processes, tools which evolution might have given us for efficiently learning those hypothesis spaces that will lead us to successful Bayesian generalizations. Such tools might include appropriately tuned stimulus metrics and topologies, as Shepard proposes, but also perhaps: unsupervised clustering algorithms that themselves exploit the size principle as defined over these metrics; a vocabulary of templates for the kinds of hypothesis spaces – continuous spaces, taxonomic trees, conjunctive feature structures – that seem to recur over and over as the basis for mental representations across many domains; and the ability to recursively compose hypothesis spaces in order to build up structures of ever-increasing complexity.

We believe that the search for universal principles of learning and generalization has only just begun with Shepard's work. The "universality, invariance, and elegance" of Shepard's exponential law (to quote from his article reprinted in this volume) are in themselves impressive, but perhaps ultimately of less significance than the spirit of rational analysis that he has pioneered as a general avenue for the discovery of perceptual-cognitive universals. Here we have shown how this line of analysis can be extended



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

to yield what may yet prove to be another universal: the size principle, which governs generalization from one or more examples of arbitrary structure. We speculate that further universal principles will result from turning our attention in the future to the interface of learning and evolution.

ACKNOWLEDGMENTS

The writing of this article was supported in part by NSF grant DBS-9021648 and a gift from Mitsubishi Electric Research Labs. The second author was supported by a Hackett Studentship.

NOTES

**1.** We derive Equation 1 as follows. Because $\mathcal{H}$ denotes an exhaustive and mutually exclusive set of possibilities, we can expand the generalization function as

$$p(y \in C|x) = \sum_{h \in \mathcal{H}} p(y \in C, h|x) \quad (12)$$

$$= \sum_{h \in \mathcal{H}} p(y \in C|h, x)p(h|x). \quad (13)$$

Note that $p(y \in C|h, x)$ is in fact independent of $x$. It is simply 1 if $y \in h$, and 0 otherwise. Thus we can rewrite Equation 13 in the form of Equation 1.

**2.** Note that in a continuous space, when $|h| < 1$, $p(x|h)$ will be greater than 1 (for $x \in h$). This occurs because $p(x|h)$ is a probability density, not a probability distribution; probability density functions may take on values greater than 1, as long as they integrate to 1 over all $x$.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

BEHAVIORAL AND BRAIN SCIENCES (2001) **24,** 641–651
*Printed in the United States of America*

# Is kinematic geometry an internalized regularity?

**Dejan Todorovič**
*Department of Psychology, University of Belgrade, Serbia, Yugoslavia*
**dtodorov@dekart.f.bg.ac.yu**

**Abstract:** A general framework for the explanation of perceptual phenomena as internalizations of external regularities was developed by R. N. Shepard. A particular example of this framework is his account of perceived curvilinear apparent motions. This paper contains a brief summary of the relevant psychophysical data, some basic kinematical considerations and examples, and several criticisms of Shepard's account. The criticisms concern the feasibility of internalization of critical motion types, the roles of simplicity and uniqueness, the contrast between classical physics and kinematic geometry, the import of perceived path curvilinearity, and the relation of perceptual and scientific knowledge.

**Keywords:** internalization of regularities; kinematic geometry; simplicity

## 1. Internalization of external regularities as a perceptual explanatory strategy

It is beyond doubt that millions of years of evolutionary processes have profoundly shaped the visual system. However, it is also beyond doubt that evolution is generally neglected by perception theorists, and that detailed accounts of its role are rather uncommon. A rare and welcome exception is the framework developed by Shepard (1984). In a nutshell, he proposes that some perceptual competencies are based on internalizations of external regularities. These regularities are characteristics of the physical environment that were more or less invariant during the evolutionary history of a species or of its ancestors. Through mechanisms of evolutionary adaptation, some outward physical facts were transformed into inward biological constraints. The organisms, having been shaped by the world, in this sense reflect its structure. In consequence, their attempts to recover that structure in the process of perception are based not only on the actual environmental circumstances or the individually acquired knowledge, but also on built-in interiorizations of exterior features.

As an example of an internalized behavioral regularity, Shepard (1984) discusses the relation of the external day-night cycle and the internal sleep-wake cycle. Why are diurnal animals active by day and resting by night? At first sight, it may appear plausible that this activity sequence is directly controlled by the environmental temporal variation of the amount of light, and that no contribution from the organism is necessary to maintain the behavioral cycle, except to be sensitive to the illumination cycle. However, experiments in which animals are kept for extended amounts of time under artificial, completely homogeneous light conditions show that in such circumstances the behavioral cycle continues with relatively small deviations, although eventually they do add up. Thus, although influenced from without, the cycle is also guided from within.

Internalization of external regularities is an intriguing and important explanatory framework for perceptual processes. However, in order to convert such a general notion into particular theories, concrete examples are needed in which the explanatory power of the approach is tested in actual applications. To date, only few cases of such explanations are available. As an example, consider a possible account of the "rigidity principle," the tendency to perceive rigid structure in geometrically ambiguous structure-from-motion displays (Johansson 1964; Ullmann 1979). One way to explain the origin of this principle is to propose that overexposure to rigid body motions in the environment somehow has induced the perceptual apparatus, over the course of evolution, to prefer the rigid interpretation over the infinity of geometrically equally appropriate nonrigid interpretations of stimuli. Another example is the "light-from-above principle," the tendency to favor those three-dimensional (3-D) interpretations of some geometrically ambiguous shaded displays which are in accord with the assumption that the represented scene is illuminated from above rather than from below (Metzger 1975; Ramachandran 1988a). That tendency may be based on the fact that the sun, the primary source of light during the evolution, was invariably located above the illuminated terrestrial scenes. Still another example is Shepard's (1992; 1994) proposal that the three-dimensionality of perceived chromatic surface color is based on the essential three-dimensionality of the color of daylight.

In this paper, I will discuss the perhaps most articulated account of this type: Shepard's theory of internalization of kinematic geometry. In the next section I will describe the phenomena that this theory attempts to explain. In the third

Dejan Todorovič is Associate Professor of Psychology at the University of Belgrade, Serbia, Yugoslavia. His fields of interest are: mathematical analyses, computer simulations, and psychophysical studies of motion, lightness, and space perception.

© 2001 Cambridge University Press 0140-525X/01 $12.50

https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

section some relevant basic kinematic facts will be recounted. In the fourth section the account of these phenomena by Shepard and coworkers will be presented. In section five I will present an examination of this account. My conclusions will be largely negative, that is, I will argue that although the notion of internalization of external regularities certainly has great merit and interest, this particular implementation of that general framework is faced with a number of conceptual and empirical problems.

## 2. Curvilinear apparent motion

In classical apparent motion displays, a figure is presented first in one spatial position and then in another. Under appropriate conditions, the two static presentations induce the impression that the figure has moved from one position to the other. The characteristics of the motion percept depend on several stimulus parameters. For the present purpose, *orientational* aspects of the stimuli are the most relevant ones. In many apparent motion studies the motion inducing figures take the form of circular disks or rings, that is, they have no intrinsic orientation. Other studies use oriented figures, but with the same orientation in the two positions, such as a vertical rectangle presented in two horizontally displaced positions. In such circumstances, one usually sees the figure move on a rectilinear path between the two positions. But what is the perceived shape of the path if the figures are presented in different orientations in the two positions? Wertheimer (1912) noted that if a line is first exposed in the vertical and then in the horizontal orientation, it is perceived to *rotate* between the two positions. More recently, several studies investigated the dependence of the shape of the perceived path on various stimulus conditions (Bundessen et al. 1983; Farrell 1983; Foster 1975b; Hecht & Proffitt 1991; McBeath & Shepard 1989; Mori 1982; Proffitt et al. 1988; Shepard 1984). Such studies involve apparent motions perceived in both 2-D and 3-D.

### 2.1. 2-D motions

Figure 1 depicts some examples of stimuli used and variables manipulated in these studies. In each of the four depicted cases, an elongated rectangle is presented in different orientations in two positions. The variables manipulated in the examples are distance, orientation difference, and symmetry. For example, the difference between Figures 1a and 1b is the distance between the centers of the rectangles. In 1c the distance is the same as in 1a, but the relative angle between the two orientations of the major axes of the rectangles is larger. In 1d the distance is the same as in 1a and 1c, and the relative orientation angle is the same as in 1a and 1b; however, in contrast to the other figures, the display is not symmetrical.

A general finding of studies of this type is that the perceived path of apparent motion is not rectilinear but *curvilinear.* The shape of the path is often assessed by having subjects indicate the perceived position of the figure (by means of a probe stimulus or an appropriate gap through which the stimulus would fit) in some intermediate location between the two displayed positions. Another general finding is that for stimuli such as Figure 1a, that is, symmetrical displays with relatively small distance and orientational difference, the perceived path has an approximately circular shape. However, with increasing distance, orientational difference, and with departures from symmetry, the perceived path, while remaining curvilinear, generally exhibits increasingly less curvature, that is, approaches rectilinear shape (McBeath & Shepard 1989; Proffitt et al. 1988).

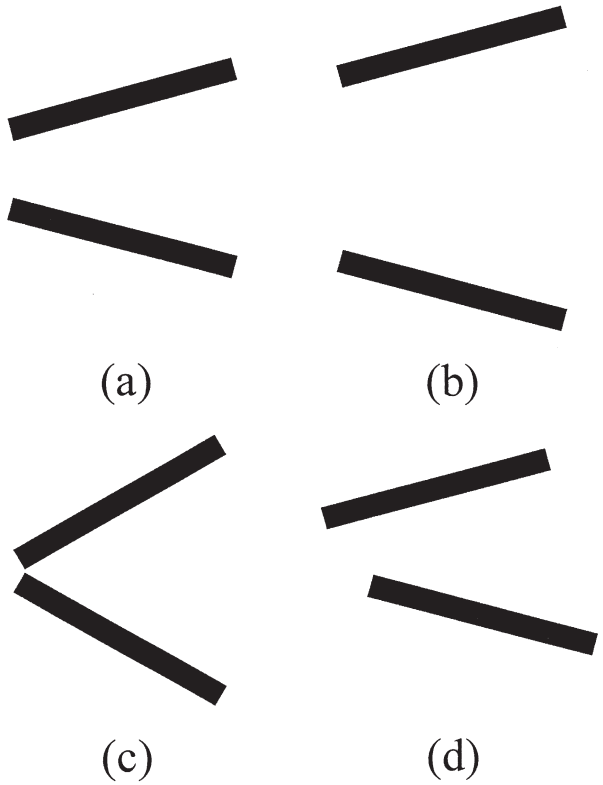



Figure 1. Examples of stimuli in 2-D planar curvilinear apparent motion studies, involving manipulations of distance, angle, and symmetry. See text for details.

### 2.2. 3-D motions

Apparent motion can also be perceived in depth. In Shepard's (1984) informally described study, the stimuli were randomized polygons presented in various orientations and sizes. The perceived motions were predominantly in depth and their motion was generally screw-like (to be described below). No use of probe stimuli was reported, so that no detailed conclusions about the shapes of the paths are available. In a study by Hecht and Proffitt (1991) the stimuli were drawings of domino-like figures, displayed in a similar fashion as in Figure 1, but depicted in different 3-D orientations. Such stimuli induced apparent motion in depth, whose path was assessed with probe stimuli. The results were similar to the studies of apparent planar motion: while in some favorable conditions the perceived path was approximately circular, in many others it fell short of 3-D circularity.

## 3. Basic kinematics

In Shepard's approach, the perceived curvilinearity in apparent motion is explained as due to internalizations of features of real motion. The study of the geometry of motion belongs to the province of *kinematics,* a branch of mechanics which is basic for physics and astronomy and has many practical ap-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

plications in engineering and robotics (Bottema & Roth 1979; Mićunovic & Kojić 1988). Thus, in order to analyze both the phenomena as well their explanations in more detail, some relevant elementary kinematic notions and facts will be recounted here. Kinematics is a large and thoroughly mathematical subject. I will only use verbal descriptions in the main text, but will illustrate them geometrically. The stress will be on 2-D motions, but 3-D motions will also be discussed.

### 3.1. 2-D motions

The issue of main interest here is the following. Let two positions of a rigid planar figure be given, denoted as I (initial) and II (final). A sequential presentation of the figure in the two positions would, under conducive conditions, induce a percept of apparent motion. But how might the figure have *really* moved in the plane from position I to position II? I will call any such transformation a "transporting motion." This section contains a survey and classification of various types of transporting motions.

Clearly, there is an infinity of possible transporting motions between any two positions. Figure 2 depicts geometric sketches of four examples of motions relevant for this discussion; Figure 3 contains additional examples in a somewhat different format, with paths calculated according to appropriate formulas, described in the Appendix. In Figure 2 positions I and II are represented as shaded forms. Two arbitrarily selected points of the figure, A and B, are indicated in their initial locations, AI and BI, and their final locations, AII and BII; point pairs such as AI and AII, and BI and BII, are referred to as "homologous" points. Each case contains, as outlined forms, several snapshots, that is, intermediate positions during the motion of the figure. The trajectory of point A is depicted as a dashed line.

In case 2a the transporting motion consists of a sequence of two phases, the translational phase and the rotational

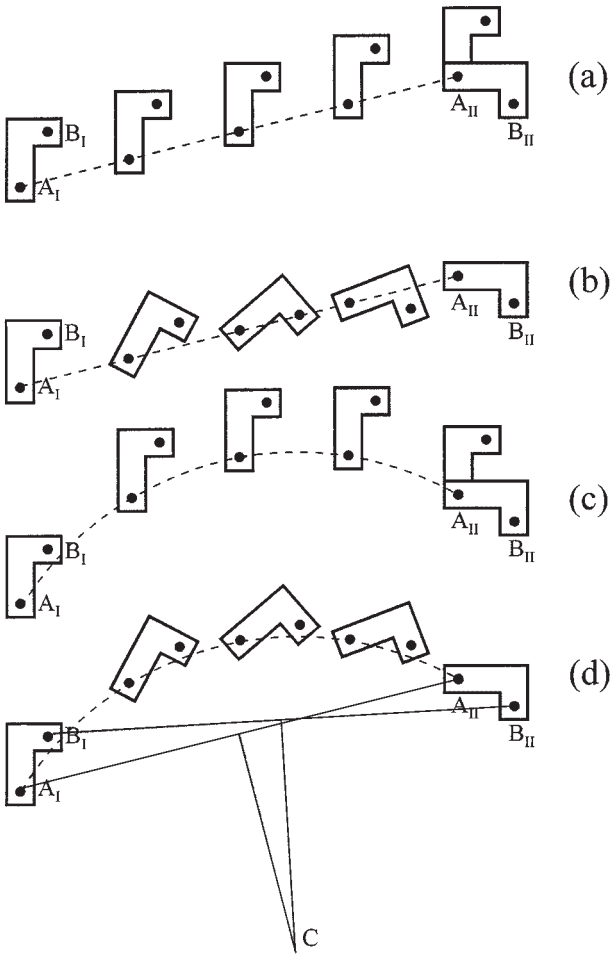



Figure 2. Examples of planar transporting motions. (a) Rectilinear translation followed by rotation about A. (b) Rectilinear translation concomitant with rotation about A. (c) Circular translation followed by rotation about A. (d) This motion can be kinematically defined in two ways. The first possibility is circular translation concomitant with rotation about (a). The second possibility is pure rotation about C. The construction of point C is explained in the text.

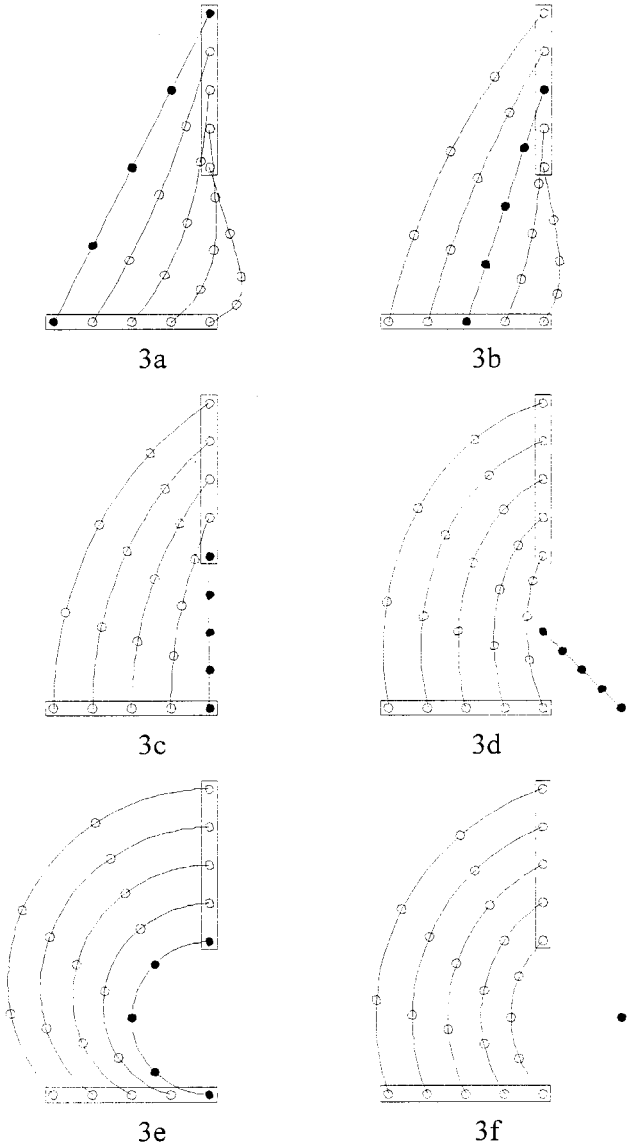



Figure 3. Examples of planar transporting motions. The black disks denote the positions of the pole. The first four figures depict rectilinear translation + rotation. (a) Leftmost point on the figure is the pole. (b) Middle point on the figure is the pole. (c) Rightmost point on the figure is the pole. (d) Pole does not belong to the figure, but is rigidly connected to it. (e) Circular translation + rotation. Rightmost point on the figure is the pole. (f) Pure rotation. The stationary pole is denoted by the black disk.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

phase. The first phase involves *rectilinear* translation, a type of motion in which the trajectories of all points of the figure are parallel straight lines of equal length. In the last position of the first phase point A has reached its final location AII, but point B has not. In the second phase, the figure rotates about A, through an angle that moves point B and all other points of the figure into their final positions. Point A is the "pole" or center of rotation. The trajectories of all points of the figure in the second phase are concentric circular segments, whose radii are given by the distances of the points from the pole A.

Several variants of this type of transporting motion can be implemented. For example, the two phases might be interchanged, such that the figure first performs the rotatory motion about AI and then the appropriate translation. Or, as represented in case 2b, the two phases might completely temporally overlap: as it translates, the figure also rotates about A, and does it in such a manner that point B arrives at location $B_{II}$ at just the moment that point A arrives at location $A_{II}$. Note that in this motion point A, the pole, travels along a rectilinear trajectory, while point B and all other points rotate about A. Thus, their trajectories are circular *relative* to the pole; relative to the stationary background the motions also have a translatory component, so that their trajectories are more complicated curves, belonging to the family of cycloids.

The choice of point A as the pole in cases 2a and 2b is arbitrary. For example, in order to perform the motion between positions I and II, point B might also serve as the pole, and would move rectilinearly from $B_I$ to $B_{II}$, while all the other points, including point A, would rotate about it. Different choices of the pole induce different trajectories of the points, but the total angle of rotation must be identical.

The effect of variation of the position of the pole is illustrated in Figure 3. This figure depicts several ways in which a rectangular object might move from a horizontal position I to a vertical position II. Five points of the figure and their trajectories are indicated. The initial and final positions of the points, as well as three intermediate snapshots, are depicted as small circles. The *filled* circles indicate the positions of the point that serves as the pole. In Figure 3a the pole is the leftmost point of the figure in the horizontal position, in Figure 3b the pole is the middle point, and in Figure 3c it is the rightmost point. Note that in these figures the pole moves on a rectilinear trajectory, different in each case, while the trajectories of the other points are curved, and are also different in different cases. In Figure 3d the translating pole does not belong to the moving figure, but is assumed to be rigidly connected to it. In this case the trajectories of all points of the figure are curvilinear. Rotation of the figure about the pole is 90° clockwise in each case.

In all these examples the translational component of the transporting motion is rectilinear. However, there is also a *curvilinear* variant of translational motion. As in rectilinear translation, in this type of motion the trajectories of all points are congruent (they have equal shape, size, and orientation, and thus are completely superimposable), but they are curvilinear. An example of such a motion, involving *circular* translation, is presented as the first phase of motion depicted in case 2c; in this phase all points move on congruent circular arcs. The second phase of the motion is a rotation about pole A, as in case 2a. Note that although trajectories of individual points are circular in both circular translation and rotation, these two types of motion are by no means identical. In rotation, all points rotate about a common center, their trajectories are concentric circular arcs whose lengths are different for different radii, and the figure changes its orientation during the motion. In contrast, in circular translation each point has a different center of rotation, but the radii are all equal, the trajectories of all points are congruent circular arcs, and the figure retains its orientation during the motion.

Case 2d presents a temporal overlap of translation and rotation, analogous to case 2b, except that in this case translation is circular: as it translates along circular arcs, the figure also rotates about point A as the pole. The role of the straight lines and point C in this figure will be discussed below. Another example of this type of motion is presented in Figure 3e, with the rightmost point as the pole.

A common feature of examples presented so far is that they are all cases of various combinations of both translations (rectilinear or circular) and rotations. Indeed, all planar motions can be represented as such combinations, with pure translations and pure rotations as special cases. Now, it may appear at first sight that when the two positions of a figure have different orientations, *both* translation and rotation would generally be necessary to transport the figure from one position to the other. Interestingly, this is not the case: it can be shown that, given two positions of arbitrarily different orientations, a specific *pure rotation* always suffices as a transporting motion, without the need of a translational component. This fact is illustrated by case 2d. Note that this case was used above to illustrate a combination of circular translation and rotation. However, this particular motion can also be instantiated by a pure rotation, but about a different, stationary center. In order to find this center a simple geometrical construction, indicated in Figure 2d, may generally be used; for exceptions, see below. It involves constructing the perpendicular bisectors of lines $A_IA_{II}$ and $B_IB_{II}$ (which join the homologous points), and finding the intersection C of these bisectors. An elementary geometric argument proves that triangles $CA_IB_I$ and $CA_{II}B_{II}$ are congruent, from which it can be deduced that the figure as a whole can indeed be transported from position I into position II by a rotation about C. Figure 3f presents another example of a pure rotational motion about a static pole, depicted as a filled circle. Note that trajectories of all points are circular arcs whose common center is the pole. As in case 2d, this particular motion can also be instantiated as a combination of translation and rotation about a moving pole, as is shown in the Appendix.

When two differently oriented positions of a figure are given, any point in the plane can be used as the pole about which the figure can be rotated from position I into a position in which it has the same *orientation* as in position II. But in almost all cases this position will be different than position II, so that an additional, translatory motion (which is orientation preserving) is necessary to complete the transporting motion. However, there is a unique point in the plane, determined by the geometric construction presented in Figure 2d, such that when the figure is rotated about it by the required angle, it is already transported to position II, and no additional translation is necessary.

The motion from I to II is not uniquely determined by the center of rotation. The example depicted in Figure 2d involves clockwise rotation, but counter-clockwise rotation


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

could also be used. Furthermore, in moving from I in either direction, the figure might have performed one or more additional full 360° rotations before coming to a stop at II. However, if one always opts for the shortest route, then there exists only a single appropriate pure rotation, except if it involves 180°, when two routes have equal lengths. Ignoring this last detail, it can be said that in this sense pure rotation is a *unique* transporting motion. In contrast, there are infinitely many *different* translation + rotation combinations, though each unique in itself, that will get the figure from I to II.

In all discussed examples positions I and II had different orientations. When they have the same orientation, a *pure translation* suffices to transport the figure from I to II, and the rotation component is not needed. In such cases the above intersection-of-perpendicular-bisectors construction fails: lines $A_IA_{II}$ and $B_IB_{II}$ are parallel, so that their perpendicular bisectors are also parallel and do not intersect, and thus no center of rotation can be constructed. This suggests that in such cases the figure cannot be transported from I to II by a pure rotation. Nevertheless, one can assert that even pure translation is in fact a rotation, but around an *infinitely* distant pole. If this is accepted, then pure rotation, just as translation + rotation, is a truly general type of transporting motion, in that it can be used to transport a figure between any two positions in the plane, regardless of their orientations.

There is another class of cases in which the above general geometric procedure fails, and that is when the two perpendicular bisectors coincide, and thus have no unique intersection point. Such a situation arises in symmetrical cases, such as those depicted in Figure 1a, 1b, and 1c. In these special cases, which are most often used in psychophysical experiments, the center of rotation can be found as the intersection of lines $A_IB_I$ and $A_{II}B_{II}$, a procedure which does not work in the general case.

In all examples presented so far, the translational component was either zero (as in pure rotation), or it had a constant direction (as in rectilinear translation), or a constantly but uniformly changing direction (as in circular translation). However, in general the direction of translation may change from point to point, so that the figure may translate along an arbitrarily curved line, while rotating concomitantly. Such motions can always be represented as pure rotations, but around a mobile pole.

The specifications of motions in the above examples are incomplete, because to define the motion of a figure it is not enough to determine the trajectories of its points. One also needs to specify the temporal manner of motion along the trajectories. The two basic possibilities are uniform and nonuniform motion. In uniform translations equal distances are covered in equal times, whereas in uniform rotations equal angles are covered in equal times. In contrast, in nonuniform motions the velocities of points may change over time. Thus, a general transporting motion from I to II may involve arbitrary accelerations, decelerations, and direction reversals.

### 3.3. 3-D motions

Kinematics in 3-D is in many ways analogous to the 2-D case, but it is also much more complex, both mathematically and visually. I will only discuss some basic facts which are relevant for the present purpose. The same main question as in the case of 2-D motions can be posed. We are given two arbitrary spatial positions of a rigid body, denoted as I and II. How might the body have moved from the initial to the final position?

As in the plane, the transporting motions in 3-D can be performed in an infinity of ways by combinations of various types of translations and rotations. Translation in 3-D is defined in the same way as in 2-D, except that the trajectories are spatial curves; rectilinear translation is specified by a constant 3-D direction. Rotation in 3-D is not defined as in 2-D with respect to a point (rotation center), but with respect to a line (rotation axis). The circular trajectories of the rotating points are not all concentric (as in the 2-D case), but they are co-axial, as their centers all belong to the rotation axis, with the planes of all trajectories being perpendicular to the axis. Similarly to the 2-D case, the temporal order of translation and rotation is arbitrary: the translation could be performed first and then the rotation; translation and rotation might be concomitant; and so on. In general, the translation direction may vary arbitrarily during the motion, and the rotation axis may change its location and orientation.

As in the planar case, there are many different translation + rotation combinations that can transport a figure from I to II. However, in contrast to the 2-D case, pure rotation in 3-D does not qualify as a general, unique type of transporting motion: given two positions of a rigid body, although in some cases it can be moved from one position into the other by a pure rotation, this is not true in general. Still, there is a particular type of a rectilinear translation + rotation combination which has a uniqueness status similar to pure rotation in 2-D. This specific combination is called "helical" or "twisting" or "screw" motion. Its specificity is that, whereas in general rectilinear translation + rotation combinations the rotation axis has a different orientation from the direction of translation, in helical motions these two orientations are equal, and thus the rotation axis is parallel to the direction of translation. It can be shown that, given two arbitrary positions of a rigid body, it can always be transported from the first position into the second position by an (almost) unique helical motion; this kinematic fact is often referred to as *Chasles's theorem.* I say "almost" unique, because the same considerations about clockwise and counter-clockwise directions and multiple turns apply as in the case of pure rotations in 2-D. The geometric construction and analytical expression of this unique transporting motion is much more complicated than for pure rotation in the 2-D case.

## 4. Shepard's account

How can the kinematic facts on possible transporting motions reviewed in section 3 help explain the empirical data on apparent motion paths described in section 2? One answer is: by applying the notion of internalization of external constraints sketched in section 1 (Carlton & Shepard 1990a; 1990b; McBeath & Shepard 1989; Shepard 1984; 1994). The general idea is that our perceptual systems have internalized some relevant regularities of real external motions, which then determine the motion impressions induced by apparent motion stimuli. In the following, the crucial arguments are presented using citations from the relevant papers:



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

> The paths that are psychologically favored . . . provide information about our internalized principles concerning the ways in which objects transform in the world. (Carlton & Shepard 1990a, p. 133)

What are these internalized principles of external motions? One possibility is as follows:

> On the basis of the assumption that we have internalized the regularities that have prevailed in the terrestrial environment throughout biological evolution, one might at first suppose that our internalized rules for the motion of rigid bodies in space would correspond to the laws of motion of classical, Newtonian mechanics. (Carlton & Shepard 1990a, p. 142)

How do the laws of classical mechanics govern object motions?

> Any continuous rigid motion would be consistent with the laws of motion in the presence of arbitrarily changing forces. In the absence of external forces (such as those of friction, air resistance, and gravity), however, the center of mass of the object must, according to classical physics, traverse a straight line at a constant velocity between the corresponding centers of the object as it appears in the two given images. (Carlton & Shepard 1990a, p. 142)

Furthermore, the appearance of apparent motion stimuli suggests an additional component:

> Because the two views also differ by a rotation, such a motion would have to be accompanied by an additional, apparent rotational transformation. (Shepard 1984, p. 425)

An example of such a type of motion, in 2-D, is presented in this paper as Figure 3b.

However, the data reviewed in section 2 suggest that such a combination of rectilinear translation and rotation is generally not perceived (but see Farrell 1983; Mori 1982). Therefore, a different source of internalized constraints is proposed, provided by the principles of kinematic geometry. Although we are able to experience any given type of *external* motion, when we are presented by discrete displays inducing *apparent* motion, "the default motions that are experienced in the absence of external support are just the ones that reveal, in their most pristine form, the internalized kinematics of the mind and, hence, provide for the possibility of an invariant psychological law" (Shepard 1994, p. 10). What is the form of this internalized kinematics of the mind? In general, what is perceived both in 2-D and in 3-D is " . . . the unique, simplest rigid motion that will carry the one view into the other . . . " (Shepard 1984, p. 426). More specifically, describing Foster's (1975b) data on perceived 2-D motions, Shepard notes that "the motion tended to be experienced over that unique circular path that rigidly carries the one figure into the other by a single rotation about point P [the pole], in the plane" (Shepard 1984, p. 425). This type of motion is depicted here in Figures 2d and 3f. Why is it that the perceptual system prefers pure rotation over rotation + translation? "The pure rotation . . . could be regarded as the simplest motion. . . . Any other motion would require, at least, a rotation through the same angle and, in addition, translation" (Carlton & Shepard 1990a, p. 152). Similar accounts are given in Shepard (1984, p. 425) and Shepard (1994, p. 8).

Analogous considerations apply to 3-D. Describing his data on perceived 3-D motions, Shepard notes that "out of the infinite set of transformational paths through which the one shape could be rigidly moved into congruence with the other, one tends to experience that unique, minimum twisting motion prescribed by kinematic geometry" (Shepard 1984, p. 425); the reference here is to Chasles's theorem. Helical motion tends to be perceived because it is "the geometrically simplest and hence, perhaps, the most quickly and easily computed. Certainly, within a general system suitable for specifying all possible rigid motions, such a motion requires the minimum number of parameters for its complete specification" (Shepard 1994, p. 7).

However, the data reviewed in section 2 indicate that in experiments pure rotation or helical motion is not generally perceived. The reported shapes of the perceived paths in these studies are mostly curvilinear, but their shape is neither purely rectilinear nor purely circular or helical. In order to account for such data, McBeath and Shepard (1989) invoke various spatiotemporal processing limitations of the visual system. On the other hand, Carlton and Shepard (1990a) argue that "our perceptual systems may have internalized still more general geometrical principles under which the physical and geometrical principles so far considered are both subsumable as special cases" (p. 168).

> The relevant principles that have been invariant throughout evolutionary history appear to be of two general types – namely physical and geometrical . . . The principles of kinematic geometry, which are in some ways more general and pervasive, may be more internalized than the principles of classical physics. Perhaps different weighted combinations of the two types of principles may yield the best accounts of data from different individuals or under different conditions. (p. 174)

Similarly, Shepard (1994) argues that internalized knowledge of classical physics "may be contaminated, to a variable degree across individuals and conditions of testing, by a more deeply internalized wisdom about kinematic geometry" (p. 8).

In summary, this account claims that during the course of evolution our perceptual systems have internalized some invariant regularities that govern the motions of external objects between two discrete spatial positions. In apparent motion displays, in which there is no real motion in the stimulus, such internalized constraints guide the percept. The discrepancies of empirical data from predictions may derive from a compromise between principles of classical physics and kinematic geometry.[1]

## 5. An examination of Shepard's account

I will discuss several issues that I find problematic for this theory. They concern questions of internalization of regularities and principles, the roles of simplicity and uniqueness, the import of perceived curvilinearity, the contrast of kinematic geometry and classical physics, and the relation of perceptual and scientific knowledge.

### 5.1. Internalization of external regularities

There is an important difference between Shepard's account of apparent motion and other evolutionary grounded explanations noted in the introduction. The difference concerns what it is that is internalized. In the examples described earlier, the presumed internalization concerns an invariant or recurrent feature of the environment, such as the light-dark cycle (in the endogenous rhythms account), the predominance of rigid bodies in the environment (in the rigidity principle account), the direction of the sun with respect to the earth (in the account of the light-from-above


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

principle), and the spectral composition of daylight (in the account of the three-dimensionality of perceived chromatic color). In all these cases, the internalization account proposes that a pervasive external regularity is the source of an evolutionary acquired internal mechanism that underlies a behavioral effect. In contrast, as argued below, in Shepard's apparent motion account such a pervasive external regularity is missing.

Consider first the idea of the internalization of classical physics. As Carlton and Shepard (1990a) themselves point out, the type of motion that is assumed to be internalized (uniform rectilinear motion of the center of mass of the body) is derived in classical physics under the assumption of "the absence of external forces (such as those of friction, air resistance, and gravity)." The problem here is that in the terrestrial environment these external forces were never absent. How then could the perceptual system have internalized this type of motion?

To illustrate this, consider a possible "classical physics world," devoid of friction, air resistance, and gravity. Accordingly, all its objects, organic and nonorganic, are perfectly smooth entities floating about in vacuum in a gravitation-free field. All they do between collisions is to translate rectilinearly while rotating concomitantly. One day the psychologists of that world perform perceptual experiments in which they project an object first in one position in space and then in another. Their subjects report that they actually see the body as moving on a rectilinear path from one position to the other. An explanation suggests itself: their perceptual systems have internalized an external regularity of their world. In such a world this would indeed be a plausible explanation. But in our world friction, air resistance, and gravity, as well as other factors are pervasive, and thus the motions of inorganic (not to mention organic) bodies are much more complicated. Thus, it is not clear what regularities the perceptual systems might have extracted during evolution.

The idea of the internalization of the principles of kinematic geometry can be criticized in a similar fashion. Consider a possible "kinematic geometry world," a universe in which the trajectories of most objects are, for some peculiar reason, predominantly circular arcs in 2-D and helical arcs in 3-D. In such a world, these types of motions might indeed have been internalized by their inhabitants. But in our world these particular types of motions do not appear to be typical or representative. Thus, there apparently is no corresponding pervasive external regularity to be internalized. Consequently, tendencies for perceived circular or helical motions can hardly be based upon internalization of invariant environmental features.

A possible reply to such criticism might be to claim that it is not the internalization of external *regularities* that is supposed to be operative in the apparent motion account, but the internalization of certain physical or geometrical *principles.* Such a formulation is indeed clearly indicated by several quotes in section 4. Note, however, that such an interpretation would concede that the apparent motion account is indeed different from other described internalization accounts in this respect. I will deal with this possibility in a later section.

### 5.2. Simplicity

An examination of Shepard's explanation of curvilinear apparent motion indicates that, in addition to internalization, other notions also play a large or even predominant role. For example, an inspection of the citations in section 4 reveals that the notion of *simplicity* figures prominently in the accounts of the shapes of perceived motions. Note that such an account does not explicitly invoke the idea of internalization of external regularities. For example, it is not claimed that pure rotations in 2-D are perceptually preferred because, say, they are internalized to a larger extent than combinations of rotations and translations. Rather, pure rotations are singled out because they are simpler. Thus Shepard's account is to an important extent based on the concept of simplicity, and not only on the notion of internalization of external regularities. Recall, however, that only in some conditions subjects report circular motions, whereas in others the perceived motion path is generally less curved. Thus, even if the propensity for this type of simplicity is in some way embodied in the perceptual system, it can easily be countered by other factors.

Now, what could be more obvious than the fact that rotation is simpler than rotation *and* translation? However, as it turns out, this matter is more complex, and the notion of comparing the simplicity of different types of motions is in fact itself not very simple or straightforward.

The intuitively compelling claim that pure rotation is simpler than rotation + translation is not only based on the fact that the former motion has one component and the latter has two, but also on the notion that "rotation" in pure rotation is the same thing as in rotation + translation. However, this is not quite the case. The two rotations do have in common the specification of the angle of rotation. However, there is a difference concerning the manner of determination of the center of rotation. In translation + rotation the pole can be chosen freely, so that any point in the plane can serve as the center of rotation. In contrast, as was shown above, in pure rotation the position of the pole must be constructed on the basis of the features of the two positions of the figure, and only a single point (in some cases infinitely distant) can serve as the pole.

There is another difference between the two types of motions, and that is, of course, that in contrast to pure rotation, in rectilinear translation + rotation two coordinates of the 2-D translation direction need also to be determined. However, the advantage of getting rid of translation is offset by the need to establish the two coordinates of the pole in pure rotation.

Furthermore, it is interesting to note that the plotting formulas that I have used to draw Figure 3 (see Appendix), are identical for cases of translation + rotation (3a–3d) and pure rotation (3f). The same program is used to draw all five cases, using different parameter values for each case. The specificity of case 3f is only in that both translation coordinates are zero. However, as zero is just another number, case 3f is just another case of motion, and, at least in such a representational format, it is not qualitatively singled out from other cases and requires the same number of parameters for its specification. In addition, as noted in section 3 and in the Appendix, the motions that were specified as pure rotations (case 2d and case 3f) can be exactly duplicated by specifying them as instances of translation + rotation.

In kinematics textbooks it is shown that general motion in the plane involves three degrees of freedom, two for translation and one for rotation. In contrast, pure rotation requires only a single degree of freedom (for the rotation angle), but only if the center of rotation can be established



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

in advance. If that is not the case, two additional pieces of information are required to specify its position in the plane. Thus, given a simplicity metric in terms of the number of degrees of freedom, translation + rotation is as simple (or as complex) as pure rotation, since the free choice of the pole does not use up degrees of freedom. Thus, by this criterion, pure rotation is not singled out from other motions by virtue of its simplicity.

In the preceding paragraphs pure rotation and rotation + translation were compared with respect to the number of pieces of information needed to specify them. As cited in section 4, Shepard (1994) suggests that simpler motions are those that are specified with fewer parameters, and that for that reason they may be calculated more quickly and easily. However, note that different parameters may themselves be computed with different ease, so that comparing just the number of informations might not be a very sensitive measure of simplicity. Therefore, rather than considering just the sheer number of parameters, it may be perceptually more relevant to discuss the potential processes by which the values of these parameters could be established.

When presented with two different spatially offset and temporally sequential positions, I and II, on what basis might the visual system come up with the parameters needed for the specification of a particular transporting motion? The angle of rotation, needed for both pure rotation and translation + rotation, may be established by the difference of orientations of the figures in the two positions. In translation + rotation, the center of rotation may be chosen to reside at some visually salient location of the figure, such as at its endpoints or center; the direction of rectilinear translation may be given by the orientation of the virtual line connecting two homologous points, which themselves may be singled out by the same visual criteria as the pole. Such motions are represented in Figures 3a, 3b, and 3c. Among them, case 3c may be singled out because it involves the shortest path of the pole.

In contrast to translation + rotation, in pure rotation the main problem is to specify the center of rotation. As noted above, the specific procedure for symmetrical displays is relatively simple, but it cannot be generalized to nonsymmetrical displays. The general procedure presented in section 3 is also simple enough to execute, once one has a compass and a ruler at one's disposal. But it is not clear how the visual system may go about to find midpoints of virtual lines connecting points that are not presented simultaneously, then to construct, at those midpoints, lines perpendicular to them (bisectors), and finally to ascertain the location of the intersection of the bisectors, which may be removed at some distance (up to infinity) from the presented figures. Perhaps the visual system could use some other, more easily computable type of construction, but this remains to be shown. Shepard (1984) suggests that the specification of the rotation center might be similar to the way this center is established in so-called Glass patterns; however, these are very different types of displays than apparent motion stimuli, and in the example pattern that is provided (his Fig. 4b), the rotation center is visually quite conspicuous. Thus this particular piece of information appears to be much harder to establish than the others.

In sum, when rotation + translation and pure rotation are compared just with respect to the number of component motions, then pure rotation (one component) appears simpler than rotation + translation (two components). However, when they are compared with respect to the number of degrees of freedom, or the number of parameters needed in a general procedure, then the two types of motions appear to be equally simple. Finally, when they are compared with respect to the easiness of visual computations of the required motion parameters, rotation + translation appears simpler. The least that can be said in conclusion is that simplicity may not be a reliable criterion by which pure rotation can be perceptually singled out from other types of motions in 2-D. A similar argument would apply concerning the presumed simplicity of helical motions in 3-D. In this case, a particular difficulty, analogous to finding the position of the pole in pure rotation in 2-D, is the procedure by which the appropriate axis of rotation can be specified in the general case, because it is complicated analytically and sometimes far from obvious.

### 5.3. Uniqueness

The notion of uniqueness features prominently in Shepard's account. The quotations cited in section 4 stress the uniqueness of pure rotation in 2-D and of helical motion in 3-D. Thus, even though these motions may not be kinematically simpler than general translation+rotation combinations, there is something that may make them kinematically salient, and that is their uniqueness. Uniqueness is clearly a different notion than simplicity (since unique entities need not be simple nor do simple entities need to be unique), nor does it relate in any obvious way to internalization of external constraints (since unique features need not be internalized nor do internalized features need to be unique). Thus, uniqueness appears to be a third, independent component in Shepard's account, besides simplicity and internalization. According to this explanatory strategy, the perceptually preferred motion paths are those that are kinematically unique.

However, the promotion of uniqueness from a kinematical feature into a perceptual principle would face some problems. For example, circular and helical motions are indeed unique in the sense explicated in section 3. But, there are other kinematical criteria in terms of which some other motions might be unique. For example, given two positions of a figure, one way to choose a particular transporting motion among the others could be to look for the motion for which some prominent feature of the figure, such as one of its endpoints, moves over the shortest path. Another candidate could be the motion for which the average length of paths of homologous points is shortest. Still another possibility would be to single out the motion for which these paths are, on average, least curved. Thus, no type of motion is *uniquely* unique: rather, uniqueness is a feature that is relative to some criterion, and the choice of one criterion rather than of another would have to be argued on some separate grounds.

But why would a perceptual system prefer a kinematically unique motion over a kinematically nonunique one in the first place? Such preference might perhaps be plausible if it were to refer to visual uniqueness or salience. However, inspect for a moment the examples of motions in Figure 3. It can be seen that all cases are different and possess some specific features. Figure 3f, kinematically unique since it is a case of pure rotation, does not appear to be especially visually unique (or, for that matter, simple). The physicist might single it out, but why should the perceiver


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

do so? Also, recall again that only in some conditions do subjects report approximately circular paths in 2-D and helical paths in 3-D. As in the case of simplicity, uniqueness – if it is embodied – can apparently be easily overridden.

### 5.4. Kinematic geometry and classical physics contrasted

According to Shepard's analysis, classical physics and kinematic geometry yield different predictions for apparent motion paths: classical physics predicts rectilinear paths, whereas kinematic geometry predicts circular and helical paths. Such an account conveys a portrayal of kinematic geometry and classical physics as two theories that can have different predictions about some aspect of reality, similar to, say, Newtonian and Einsteinian theory providing mutually incompatible accounts of the speed of light or the shape of its trajectory.

However, such a portrayal is not appropriate. Classical physics *adds* to kinematic geometry, it does not contradict it. Physics textbooks often treat motion by first discussing the more mathematical aspects (kinematics), and then introducing the more physical notions (dynamics). Whereas kinematic geometry describes the ways bodies move, classical physics, accepting this description, goes on to inquire about the physical causes of their motions. Thus kinematic geometry involves the concepts of displacement, trajectory, velocity, and so on, and classical physics involves, in addition, the concepts of mass, force, inertia, and so on. For example, kinematic geometry *describes* the shapes and velocities of trajectories of heavenly bodies, whereas classical physics deduces these shapes and velocities, using a richer set of concepts and laws.

With respect to the issue at hand, a rectilinear path is as compatible with kinematic geometry as is a circular path. Kinematic geometry does not "prescribe" a circular or helical path, since it is not in the business of prescribing paths but describing them. It is classical physics that, given some additional assumptions, singles out the rectilinear path among the many possible paths offered by kinematic geometry. Thus, the classical physics prediction is not a rival of the kinematic geometry prediction, and a test of perceived path shape is not a test between kinematic geometry and classical physics.

### 5.5. The import of path curvilinearity

The main empirical support for Shepard's account is the curvilinearity of perceived paths in apparent motion displays. However, curvilinear paths are not exclusively indicative of circular motions (in 2-D) or helical motions (in 3-D). In fact, they are just as compatible with many translation + rotation combinations. As illustrated by examples in Figure 3, all 2-D motions (except for pure rectilinear translation), involve curvilinear translations for all figure points, except for rectilinearly moving poles. The geometrical center of the figure moves rectilinearly only if it is the center of rotation (Fig. 3b), otherwise it moves curvilinearly. Several cases of translation + rotation combinations (such as Figs. 3c, 3d, and 3e) appear visually relatively similar to pure rotation (Fig. 3f), and it might be difficult to distinguish them on phenomenal grounds, or to decide to which category a given apparent motion trajectory belongs. Furthermore, some empirical studies have provided evidence that perceived paths in some apparent motion displays are not purely circular but are rather translation + rotation combinations (Farrell 1983; Mori 1982). In sum, the curvilinearity of perceived motion paths in 2-D does not provide much support for the idea that a crucial component of their shape is best described by pure rotation. Similarly, there appears to be no specific empirical evidence that perceived paths in 3-D are generally preferentially truly helical (meaning that the translation direction must be parallel to the rotation axis) rather than general translation + rotation combinations (in which this parallelism does not necessarily have to hold).

### 5.6. Perceptual and scientific knowledge

It was argued above that it is not likely that the visual system has internalized the particular *types* of motions that are central in Shepard's account. However, one could argue that what was internalized instead are some kinematic *principles* of motion, such as Chasles's theorem. The main problem for this idea is to make plausible how and why the visual system would have internalized such principles. They are not "out there" in the same sense as external regularities, which are recordable by physical or biological sensory systems. Rather, these principles involve some mathematical truths concerning the generality and uniqueness of a specific class of potential transporting motions between two arbitrary discrete positions. This type of issue is of great interest to mathematicians and is therefore duly discussed, among other theorems, in kinematic textbooks. But what would impel a biological visual system to pose, let alone solve such a problem? As noted, such motion types appear not to be predominantly encountered in the natural environment. In technical applications they are used in some cases but not in others, and the decision which concrete type of motion is chosen in a particular mechanical device is not governed by mathematical uniqueness but by technical efficiency.

Consider the general idea that, through their contact with the environment, our perceptual systems have extracted certain principles that are strongly analogous to laws, axioms, and theorems articulated by the scientific community. Such a notion is intriguing, but it should be treated with caution, and a more concrete analogy between the operation of the perceptual system and the conduct of scientific inquiry may be grossly inappropriate. For example, it is certainly true that biological organisms have acquired immense experience of geometric features of their environment. But does that mean that their visual systems must embody and apply the basic axioms and theorems one would find listed in geometry textbooks such as, say, that the square on the hypotenuse is equal to the sum of the squares on the kathetes? Or, note that all our life we are exposed to and are ourselves the sources of physical forces of various kinds. But can we conclude from this that our visual system must have extracted the overarching principle that Force is equal to Mass times Acceleration, and that it applies this principle in the interpretation of visual events?

The task of the visual system is, in part, to inform the perceiver about the makeup of the current environment. The task of the scientist is, in part, to extract analytic order out of bewildering complexity. These two tasks are different, and the ways they are accomplished may well also be different. Furthermore, the type of information offered by vision



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

may not always be conducive to the sort of problems faced by science. This means that when we theorize about the world we often have to think beyond what we see. The history of the physics of motion provides an example of adverse effects that everyday experiences may have on the development of scientific theories. Aristotelian physics assumed, quite plausibly, that whenever a body is in motion, some force must be at work, and that with the cessation of the force the body must stop. After all, what is more obvious than the fact that if you want to move a rock you must push, and if you stop pushing, it will stop moving. Thus, in Christian cosmology angels had to be recruited to eternally push the planets along crystalline spheres to prevent them from stopping. It took the brilliance of a Galileo to conceive the concept of inertia which claims, against all the evidence of the terrestrial senses, that once in motion a body would forever move on, without the need of any force to keep it going, provided that no forces act on it. To make this plausible, we can imagine a perfectly smooth horizontal surface on which a perfectly smooth ball is set rolling. Then we realize that there is nothing to make it stop. But this is a *thought* experiment. No one has ever *seen* such perfect objects. It took the genius of a Newton to establish that when forces are involved it is not in motions as such but in changes of motions. Thus a real ball on a real surface will eventually stop rolling, not because some force has ceased to act, but because another force, friction, acts against inertia.

It is not hard to find additional examples in which our individual senses are likely to delude our common sense about some aspect of reality, and where scientific progress is achieved only through increasing abstractions from the sensory givens. Such examples indicate that although perceivers and scientists may share the task of extracting regularities, their data, competencies, strategies, and goals are not the same. This makes the task of those perceivers who are also scientists of perceivers all the more difficult.

## 6. Conclusion

What is the relation between empirical data on curvilinear apparent motion and the particular circular/helical motions described in kinematic geometry? Why is it that stimuli such as those depicted in Figure 1 induce the types of apparent motions described in section 2? According to the argument presented here, it is not because such motion types are internalized, since they are not predominant in the exterior in the first place. It is not because they are kinematically simpler than other motions, since it is questionable that they are kinematically simpler. They *are* kinematically unique in a specific sense, but there are other types of uniqueness as well, and it is not clear why a perceptual system should care about kinematic uniqueness anyway. It is not because principles of kinematic geometry are preferred over principles of classical physics, since these two fields are not in predictive rivalry. It is not because empirical data predominantly indicate such motion types, since they are compatible with many other types of motion as well. Finally, it is not because certain principles of motion are extracted by the visual system from the environment, as it is hard to see how and why they should be internalized.

It should be stressed that these negative conclusions do not apply to the general idea of the internalization of external regularities. It is indeed very plausible that perceptual mechanisms are affected by evolutionary adaptational processes of the species in its contact with the environment. However, the unpacking and empirical testing of this notion is by no means straightforward, and remains as an important issue for further research.

## APPENDIX

Motion of a rigid body can be decomposed into translation and rotation. Symbolically, M = T + R. This decomposition can be mathematically represented in several forms, usually involving matrices and/or vectors. Here I will use the coordinate (parametric) form, which is less elegant but is more suited for plotting. It involves separate equations for the two Cartesian coordinates, that is, $M_x = T_x + R_x$, and $M_y = T_y + R_y$.

The rotation components are identical for all six examples presented in Figure 3. They take the form:

$$R_x(t) = P_x + (A_x - P_x)\cos\omega t - (A_y - P_y)\sin\omega t$$

$$R_y(t) = P_y + (A_x - P_x)\sin\omega t + (A_y - P_y)\cos\omega t$$

Here $P_x$ and $P_y$ are the coordinates of the pole (rotation center), and $\omega$ is the angular velocity, assumed as constant. These three parameters specify the rotation component. Time is denoted as t, and $A_x$ and $A_y$ are coordinates of points on the moving body (or rigidly attached to it) in the initial (horizontal) position. Five representative points are depicted in the examples. The origin of the Cartesian coordinate system was chosen to coincide with the leftmost point. Thus, in the initial (horizontal) position, I, all points lie on the x-axis, with the x-coordinates being 0, 0.25, 0.50, 0.75, and 1, and the y-coordinates being all zero. In the final (vertical) position, II, all points have 1 as the x-coordinate, and the corresponding y-coordinates are, respectively, 2, 1.75, 1.50, 1.25, and 1.

The translation components for cases 3a–3d and 3f take the form:

$$T_x(t) = D_x\, t$$

$$T_y(t) = D_y\, t$$

Here $D_x$ and $D_y$ give the direction of the translation, which is assumed constant.

Note that for t = 0, $M_x(t) = A_x$, and $M_y(t) = A_y$, that is, in the beginning the points are in their initial positions. As time increases, $M_x(t)$ and $M_y(t)$ change, that is, the points move, and their trajectories are different for different values of $P_x$, $P_y$, 4, $D_x$, and $D_y$. However, for t = 1 all points end up in final positions which are identical in all six cases. The particular values of these five constants for cases 3a–3d and 3f are given in Table 1.

Table 1. *Parameters for motion equations corresponding to examples depicted in Figure 3*

| CASE | $P_x$ | $P_y$ | $\omega$ | $D_x$ | $D_y$ |
|---|---|---|---|---|---|
| 3a | 0 | 0 | $-\pi/2$ | 1 | 2 |
| 3b | 0.5 | 0 | $-\pi/2$ | 0.5 | 1.5 |
| 3c | 1 | 0 | $-\pi/2$ | 0 | 1 |
| 3d | 1.5 | 0 | $-\pi/2$ | −0.5 | 0.5 |
| 3f | 1.5 | 0.5 | $-\pi/2$ | 0 | 0 |

It can be seen that in all cases the extent of rotation is $\pi/2$ clockwise. In cases involving translation+rotation (3a–3d), the rotation pole in the initial position is chosen to lie on the x-axis ($P_y$=0 in all cases). The coordinates of the direction of translation are given as the differences of the coordinates of the pole in the final and in the initial position. In the case involving pure rotation (3f) the translation coordinates are zero. The coordinates of the pole in this case can be found through the geometric construction de-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

scribed in the text, or by corresponding analytical formulas (Bottema & Roth 1979), given by

$$P_x = [O_x - O_y \cot(\omega/2)]/2$$

$$P_y = [O_x \cot(\omega/2) + O_y]/2$$

Here $O_x$ and $O_y$ are the coordinates of the point to which the origin of the coordinate system is transported by the rotation, and $\omega$ is the angular extent of rotation. In the case at hand, $O_x = 1$, $O_y = 2$, and $\omega = -\pi/2$, so that the coordinates of the pole are $P_x = 1.5$, $P_y = 0.5$, as given in Table 1.

In case 3e, the translation is not rectilinear but has the shape of an arc of a circle. Its components take the form:

$$T_x = R \cos(\sigma + \tau t) - R \cos \sigma$$

$$Ty = R \sin(\sigma + \tau t) - R \sin \sigma$$

Here R is the radius of the circular arc, $\sigma$ is the angle of the starting point of the arc, and $\tau$ is its angular extent. In the present case $R = 0.5$, $\sigma = -\pi/2$, and $\tau = -\pi$. The rotation parameters are $P_x = 1$, $P_y = 0$, $\omega = -\pi/2$. Thus the pole is the same as in case 3c, but in this case its path is semicircular.

The motion in case 3f, which was represented above as an instance of pure rotation, can also be defined as a case of circular translation + rotation, with the following parameters. As in case 3c, the coordinates of the rotation pole are $P_x = 1$, $P_y = 0$, and the extent of rotation is $\omega = -\pi/2$. However, the parameters of circular translation in this case are $R = \sqrt{0.5}$, $\sigma = -3\pi/4$, and $\tau = -\pi/2$.

## NOTE

**1.** The scientific fields that Shepard calls "kinematic geometry" and "classical physics" correspond to topics often referred to in mechanics texts as "kinematics" and "dynamics," but I stick with Shepard's labels in this paper.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

# Open Peer Commentary

*Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this topic. Integrative overviews and syntheses are especially encouraged.*

## Colour generalisation by domestic chicks

R. Baddeley, D. Osorio, and C. D. Jones

*School of Biological Sciences, University of Sussex, Brighton BN1 9QG, United Kingdom.* **rjb@biols.sussx.ac.uk d.osorio@sussex.ac.uk c.d.jones@sussex.ac.uk**
**http://www.biols.susx.ac.uk/Home/Roland_Baddeley/**

**Abstract:** We present data on colour generalisation by chicks relevant to Tenenbaum and Griffiths' (T&G) Bayesian framework. Chicks were trained with either one or two colours, and tested for interpolation and extrapolation. T&G's framework predicts an observed lack of extrapolation on the red to yellow line in colour space. A modification incorporating stimulus uncertainty deals with a prototype effect, where an intermediate is preferred to exemplars. After training to complementary colours, chicks do not generalise across an intermediate grey as T&G predict.
**[TENENBAUM & GRIFFITHS]**

**TENENBAUM & GRIFFITHS' (T&G**'s) framework for classification on single dimensions is clear and elegant. Essentially the proposal is to integrate (average) over *all* plausible interpretations consistent with the data. Taking an experimental approach to this questions we (Jones et al. 2001) have studied colour generalisation by domestic chicks. Subjects were trained to either one or two example colours, chosen according to their locations in a chicken cone-photoreceptor-based colour space (note that colour names used here are simply for the convenience of human readers, and do not reflect any assumptions about chick colour perception). Generalisation was tested for evidence of extrapolation and interpolation, with the two training colours separated by an approximately set number of just noticeable distances.

For the "red to green" line in colour space, the observed extrapolation fitted the predictions of **T&G**'s model: values more extreme than the two example colours were no more frequently selected after training with two examples (point *c* in Fig. 1), than with one (point *b* in Fig 1.). By comparison, extrapolation is predicted by some parametric Bayesian models that attempt to find the best fitting distribution for the data (such as a Gaussian one; Fried & Holyoak 1984). Also the basic interpolation prediction of the **T&G** model is supported: colours between the two examples were chosen more than predicted by the sum of responses to each colour trained individually (in Fig. 1, the value of point *e* exceeds the sum of *d* plus the equivalent value for chicks trained solely to yellow; see Jones et al. 2001).

One aspect of the interpolation was at variance with the specifics of the **T&G** proposal. This was: the evidence of a "prototype" effect, where untrained intermediate colours (Fig. 1 point *e*) were preferred to either of the trained colours (points *a, f*). Although inconsistent with the **T&G** proposal, only a simple modification is required to cope with this result. This recognises that there is uncertainty associated with any colour measurement, because of photoreceptor noise and variable illumination. A full Bayesian approach should integrate over this uncertainty. In contrast to a concept such as number, where one example is unlikely to be mistaken for another, in colour vision there is considerable uncertainty in the stimulus value. By integrating over this uncertainty (say, modelled by a Gaussian distribution in colour space) – and given the basic categorisation mechanism proposed by **T&G** – intermediate colours will indeed be more likely examples of the stimuli than the trained colours themselves. This is because the trained colours lie at the ends of the range of positive stimuli, and therefore may be confused with stimuli lying beyond this range. There is far less chance that an intermediate colour could be confused in this way.

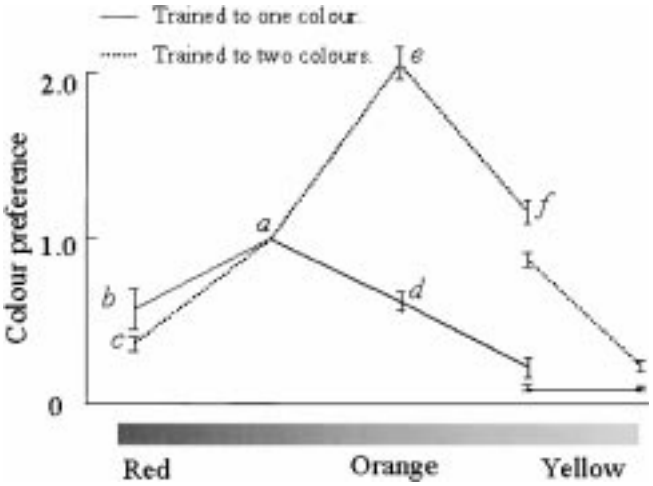



Figure 1 (Baddeley et al.). Experimental results showing preferences of chicks trained to either one (red) or two (red and yellow) example colours. Results compare preferences of the red to novel colours that were either intermediate between the two exemplars in a chick colour space (interpolation), or lay beyond the range of the two training colours (extrapolation). Responses are normalised to the preference for the red colour rewarded in both training conditions; error bars give the SEM of the preference compared to the red ($n = 10$ sets of chicks).

As stated, with a minor modification, the classification of colours in the red-green direction can be accounted for within the **T&G** framework. Similar effects are seen for interpolation between blue and green across turquoise, and between blue and red across purple. In contrast, no interpolation (or extrapolation) was found for generalisation between the complementary colours, yellow and blue, across grey. This was not because these colours were more discriminable, but because they were equally separated in terms of just noticeable differences. Such observations are difficult to fit into any general theory of classification.

## Generality, mathematical elegance, and evolution of numerical/object identity

Felice L. Bedford

*Department of Psychology and Program in Cognitive Science, University of Arizona, Tucson, AZ 85721.* **bedford@u.arizona.edu**

**Abstract:** Object identity, the apprehension that two glimpses refer to the same object, is offered as an example of combining generality, mathematics, and evolution. We argue that it applies to glimpses in time (apparent motion), modality (ventriloquism), and space (Gestalt grouping); that it has a mathematically elegant solution of nested geometries (Euclidean, Similarity, Affine, Projective, Topology); and that it is evolutionarily sound despite our Euclidean world.
**[SHEPARD]**

Not since Hemholtz has the vision for vision been so grand. First, note that **SHEPARD**'s approach points the way to a long overdue task: to compile what can be considered evolutionary constraints on perception. A few of those constraints are: sunlight comes from above (Ramachandran 1988b); one object cannot be in two locations at the same time (Bedford 1999); babies know they can fall down and not up (visual cliff; Gibson & Walk 1960); black dots can


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

only match black dots in stereopsis (Marr 1982); and Gogel's "specific distance tendency" maintaining that in the absence of definitive depth cues, an object will appear to be two meters away (Gogel 1972). These constraints vary from specific to general. Compiling, sorting, weeding, reducing will lead to a set of core constraints on how we perceive and reason about the world.

But the focus here is on **SHEPARD**'s goal of integrating generality, mathematical elegance, and evolution, by providing an example inspired by SHEPARD's long-standing broad goals (Bedford, in press). Numerical or object identity refers to perceiving and knowing when an object seen at different times refers to different glimpses of the same object. If a rock is thrown behind a dense bush, we usually believe that it is the same rock which emerges out the other side, not that there are two rocks, one of which remains behind the bush.

***Generality.*** I suggest the object identity problem is broader. In addition to glimpses separated in time, a decision is required when the glimpses are separated by modality. Suppose you are looking at your pen while writing a note. How do you know that the pen you are seeing and the pen you are feeling refer to one and the same pen? You see the lips of a ventriloquist's dummy moving and hear the voice of the ventriloquist. You perceive both the sound and the sight – in this case erroneously – as coming from the same object, the dummy. Suppose the samples occur at the same time *and* the same modality, but differ only in spatial location, for example, two tennis balls a few inches apart. Though usually regarded as definitive for two objects, the object identity decision is required here as well. The two samples could be produced by a single tennis ball viewed with diploma ("double vision") or through a mirror, or could even be a dumbbell properly thought of as a single object. The Gestalt grouping principles are descriptive rules for determining when spatially separated samples belong to the same object, that is, object identity. Object identity applies to different times, different modalities, different spatial locations, and even different eyes ("correspondence problem"). It forms the basis of such diverse phenomena as apparent motion, ventriloquism, prism adaptation, Gestalt grouping, priming, and stereopsis, which in turn reflect everyday accomplishments.

***Mathematical elegance.*** Abstracting away from content, the question in its most general form is: how are two samples determined to refer to the same object or to different ones? Following **SHEPARD** and Dennett (1996), problems with the same formal structure suggest a common solution. In nature, samples have extended contours, that is, forms; geometry, the study of form, is a natural candidate. The solution involves a whole set of nested geometries that fit inside one another like Russian nesting dolls; the familiar Euclidean geometry is only the beginning, the smallest "doll" within the set. Felix Klein (1893/1957) showed how different transformations produce different geometries of increasingly larger size within which more and more forms are equivalent. In Euclidean geometry, a square and a displaced square (isometric transformation) are identical, but if the next most radical change to a square is permitted such that it can be stretched uniformly (similarity transformation), this gives rise to a slightly bigger geometry, Similarity geometry, within which a square, a displaced square, and squares of different sizes are all considered the identical form. Next in the hierarchy is Affine geometry, which adds rectangles and sheared squares to the equivalence class, followed by Projective geometry, which broadens to include trapezoids, and finally, Topology, which is produced by such radical transformations that squares and circles are also equivalent.

For object identity, the more radical the transformation between the samples, the less likely that they will be judged as originating from the same object. When there are multiple samples, a mate for a sample will be chosen from the lowest level of the hierarchy available. For instance, in apparent motion, if there is a choice between seeing a square move to another square or to a rectangle, object identity will favor the square, but if the choice is between a circle and a rectangle, now object identity will favor the rectangle. Thus, the hierarchy has the desired property that the identical two stimuli will sometimes be judged to refer to the same object, but sometimes not. Transformations from isometric to topological span a range wide enough to apply to nearly all transformations encountered from rocks thrown behind bushes (isometric) to crumpled clothing (topological), as well as image transformations that result from our own moving, tilting, and twisting. It is a mathematical solution that has breadth, generality, and elegance.

***Evolution.*** However, doesn't this solution violate **SHEPARD**'s entire thesis that only conditions that prevail in the world will be internalized? As he notes, the space in which we evolved is three-dimensional and Euclidean; yet the above theory uses many geometries that are not Euclidean. Interestingly, there is no contradiction. All the rules of Euclidean geometry can be derived from Euclid's original axioms and postulates. Within this axiomatic approach, removal of axioms produces the more general geometries. For instance, removal of the postulate on angle enables Affine geometry. As Cheng and Gallistel (1984) argue, natural selection would not favor getting an assumption wrong but could fail to capture all the available principles.

An ingenious evolutionary solution may allow observers to jump between the geometries by alternately giving up and gaining assumptions as the situation warrants. As **SHEPARD** argues, in physics, problems have been formulated and reformulated before obtaining generality. I believe removing the restriction of Euclidean geometry is the right reformulation. I am also convinced **SHEPARD** would agree.

ACKNOWLEDGMENTS

Thanks to William H. Ittelson, Jason Barker, and Christine Mahoney for helpful discussion. This work was supported by a grant award from the Office of the Vice President for Research at the University of Arizona, funded by the University of Arizona foundation.

## If a tree falls in the forest and there is nobody around, does Chasles' theorem still apply?

Marco Bertamini

*Department of Psychology, University of Liverpool, Liverpool, L69 7ZA, United Kingdom.* **m.bertamini@liverpool.ac.uk**
**http://www.liv.ac.uk/Psychology/VP/**

**Abstract:** The limitations of the concept of internalised kinematic geometry have been recognised by Barlow, Hecht, Kubovy & Epstein, and Todorovič. I am in agreement but I still find the perception of curvature in two frames of apparent motion fascinating and I suggest some new directions.

**[BARLOW; HECHT; KUBOVY & EPSTEIN; SHEPARD; TODOROVIČ]**

**BARLOW, HECHT, KUBOVY & EPSTEIN,** and **TODOROVIČ** all argue against internalisation of kinematic geometry. I have argued along these lines in the past (Bertamini 1996; Bertamini & Smit 1998); in particular, I share TODOROVIČ's opinion that there is a fundamental difference between internalisation of physics and internalisation of mathematics (i.e., kinematic geometry). A succinct way of summarising the issue is as follows: all theories of perceptual phenomena can be cast in mathematical terms; when we need to choose between them, the differences – apart from their predictive power – can only be in their elegance or in their simplicity. Elegance brings the observer back into the equation and is therefore incompatible with internalisation, whereas simplicity is a principle closely related to Occam's razor. The only possible support for internalisation of geometry is therefore simplicity. But as TODOROVIČ shows, it is not always possible to establish which solution is simpler, because it depends on how the problem is formalised.

I also have more apparent motion data to contribute. Following the work of Proffitt et al. (1988) and using the window technique (McBeath & Shepard 1989), we investigated systematically the effect of alpha. Figure 1a shows two frames (the first is the one on



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

the left), alpha is −8 deg, and the grey circle shows the rigid rotation. We found that the perceived curvature increases monotonically as alpha goes from −45 to 45 (Bertamini & Smit 1998). Note that changes in alpha do not change the center of rotation, and that this effect is therefore inconsistent with the idea of internalised kinematic geometry. It suggests instead that motion tends to be orthogonal to the orientation of the object in the first frame.

Some observations on symmetric shapes are also important. In the case of a solid rectangle (Fig. 1b) there are not only the two paths that are always possible on the basis of Chasles' theorem (90 and 270 deg, respectively). There are two more paths identical to the first two except that alpha differs by 90 deg. Ignoring the longer paths, we still have a conflict between two possible solutions with the same angle of rotation. Both solutions can be seen, but the motion orthogonal to the orientation of the object (alpha=0) is seen more often by naive subjects. In this solution the object extends farther from the center of rotation, therefore we may be observing a difference in torque. If so, this would be an effect related to the physics of the event, not its kinematic geometry.

More axes of symmetry can be present in an object, such as in the case of an equilateral triangle (Fig. 1c). Kolers and Pomerantz (1971) have noticed that both rotation in the plane and rotation in depth can be seen in such cases (the depth solution being more likely when there is longer presentation time). What is important here is that a depth rotation of 180 deg is seen at least as often as a rotation in the plane of 60 deg. Surely this is a problem for an argument based on the simplicity of motion. We went even farther and tried quasi-symmetrical stimuli (Fig. 1d). Remarkably, motion in depth is seen even when the 60 deg rotation is a rigid motion, whereas a rotation in depth of 180 deg entails a shape change (one arm getting longer as the object moves).

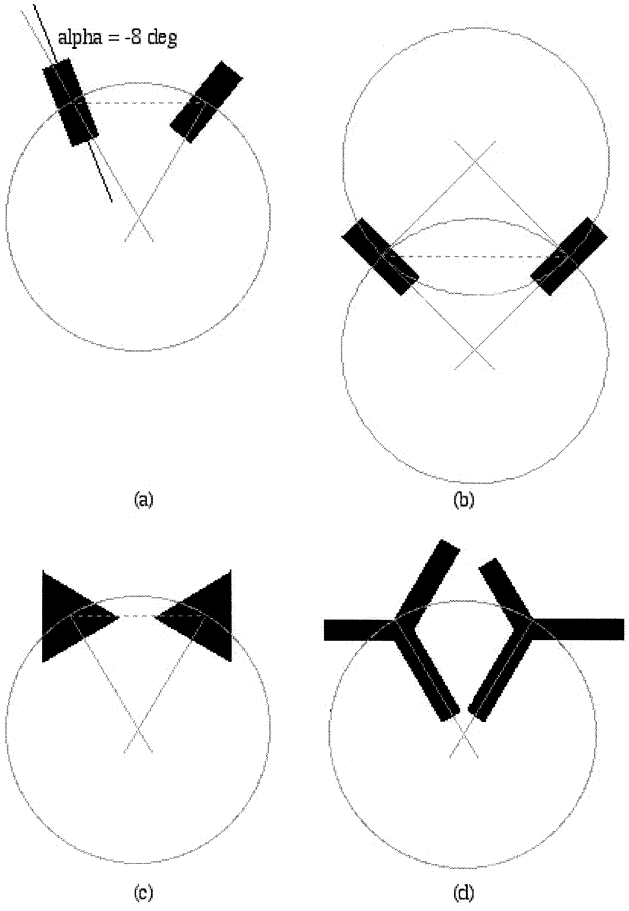



Figure 1 (Bertamini). To see the animations go to: http://www.liv.ac.uk/~marcob/todorovic.html

**SHEPARD** discusses the case where motion in depth by 180 deg is preferred to rotation by 180 deg and suggests that the reason depth rotation is preferred is because it is more consistent with retinal stimulation (i.e., if motion in the plane had taken place it would have been detected). I doubt that this could account for the case where 180 deg is compared to 60 deg, but it is a useful way of looking at the problem (i.e., what is the most likely motion given the available evidence), especially if we agree that apparent motion is a solution to poor temporal sampling (Watson & Ahumada 1983).

I hasten to acknowledge the noisiness of these data. Everybody looking at these displays will notice their inherent ambiguity. Percepts can and do change even for one individual over time. This multistability needs to be taken into account in any theory. I suggest that this multistability could be used constructively to study the way shape is represented. Taking the example of the equilateral triangle, the three axes of bilateral symmetry are identical from a geometrical point of view, but perceptually they are not. At any one time, one vertex is seen as the top and the opposite side as the base of the object. Such a chosen axis of orientation is important in determining the motion of the object (for other effects of shape on apparent motion, see McBeath et al. 1992). When the equilateral triangle is seen as oriented horizontally, the motion in depth (but not the rotation in the plane) is around a pivotal point along the axis of elongation. The importance of axes in constraining perceived motion is consistent with what **SHEPARD** is arguing, except that this does not mean that kinematic geometry has been internalised, it means that the representation of shape is not independent from the representation of motion.

We have recently found effects of pivot points in how motion is perceived, using random dot configurations (Bertamini & Proffitt 2000). These are all examples where the system assumes (or infers from spatial information) mechanical constraints on motion (Hoffman & Bennett 1986). Given the environment in which we live, this may be the best strategy. The only case in which mechanical constraints to motion do not exist are particle motions and they are not as common as extended object motion and joint motion. **SHEPARD** claims that physics would predict straight paths for the center of mass, but although this is true it misses out on the fact that given a certain shape not all motions are equally likely. Probably no tree that has ever fallen in a forest moves along a straight path, instead it rotates around a center at the base of its elongation.

Finally, as an aside, let me point out that the preference for motion orthogonal to object orientation is not a general effect. Werkhoven et al. (1990) have found quite the opposite result in short term apparent motion. This difference may be related to how the aperture problem affects the system at different scales. But this is another story.

## "First, we assume a spherical cow . . . "

Lera Boroditsky[a] and Michael Ramscar[b]

[a]*Psychology Department, Stanford University, Stanford, CA, 94305–2130, USA;* [b]*School of Cognitive Science, University of Edinburgh, Edinburgh, EH8 9LW, Scotland.* **lera@psych.stanford.edu**
**http://www-psych.stanford.edu/~lera/michael@dai.ed.ac.uk**
**http://www.dai.ed.ac.uk/homes/michael/**

**Abstract:** There is an old joke about a theoretical physicist who was charged with figuring out how to increase the milk production of cows. Although many farmers, biologists, and psychologists had tried and failed to solve the problem before him, the physicist had no trouble coming up with a solution on the spot. "First," he began, "we assume a spherical cow . . . " **[TENENBAUM & GRIFFITHS]**

**TENENBAUM & GRIFFITHS** (henceforth **T&G**) present an ambitious attempt at a computational framework encompassing generalization, similarity, and categorization. Although it would seem elegant to account for all of similarity and/or categorization in a simple unitary framework, the phenomena in question are almost certainly far


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

too complex and heterogeneous to allow this. A framework this general will inevitably fail to capture much of the intricacy and sophistication of human conceptual processing. That is, it may turn out to be a theory about spherical cows rather than cow-shaped ones.

**T&G** propose a model of similarity as generalization based on Bayesian inference. However, although T&G specify a framework (essentially, Bayes' rule and some ancillary equations), they fail to specify a procedure for generating, weighting, or constraining any of the input into this framework. At times, T&G base the representations in their hypothesis space on people's similarity judgments. It is hardly surprising that a model with people's similarity judgments built in can compute similarity. Further, the basis for T&G's claim that similarity is based on Bayesian generalization becomes unclear – in their model, generalization appears to be based on similarity and not the other way around. At present the framework relies solely on hand-coded and hand-tailored representations, while the few predictions it does make (relying on asymmetrical comparison and the size principle) are not borne out by data. We review just a few of the complications as illustrations below.

People's similarity judgments are based on a myriad of contextual, perceptual, and conceptual factors. In carrying out a comparison, people need to choose a way to represent the things to be compared as well as a strategy for comparing them. This means that a comparison between the same two items in different circumstances will yield different results. For example, in a replication of **T&G**'s study shown in the left panel of Figure 6 (with right-left position counterbalanced), 62% of our subjects picked the object-match (a) as most similar to the top example. But, if subjects were first given the example shown in the right panel of Figure 6 and then the question in the left panel, then only 33% picked the object-match. Changing how likely it was for people to notice and represent the relational structure of the stimuli had a dramatic effect on the results of the comparison. In another example, subjects were asked to say which of AXX or QJN was most similar to AHM (a problem structurally similar to T&G's in Fig. 6), and 43% chose QJN when the letters were presented in Chicago font (which makes all the letters look boxy). When the same letters were presented in Times font (which emphasized the pointy ends of the A's), only 17% chose QJN. Thus, even a trivial change in the perceptual properties of the stimuli can have a dramatic effect on how people choose to represent and compare the arrays.

Nothing inherent in **T&G**'s framework predicts these kinds of results. Although T&G's framework might allow for perceptual similarity, effects of context, and other factors to be coded into the hypothesis space, it is disappointing that it is these back-door (i.e., coded-in and not necessarily principled) elements, and not anything about the framework itself, that carry all of the explanatory power. Moreover, at times the specifics of the framework can even prevent the back-door solutions from working, even when these solutions are probably the psychologically correct ones. Consider the following example: When subjects were asked which of 1-911-ANALOGY or 1-208-BKSDEMG was most similar to 1-615-QFRLOWY, 75% of the subjects chose 1-208-BKSDEMG (chi^2=5.00, $p < .05$) even though 1-911-ANALOGY shares 4 extra features with the base example, and the "1 in position 3, L in position 8, O in position 9, and Y in position 11" hypothesis is more than 72,000 times more restrictive than the "all different letters" hypothesis. Despite an advantage of more than 72,000 to 1, the size principle proposed by **T&G** as a new universal had no effect. We doubt that any one of our subjects even considered the "1 in position 3, L in position 8, O in position 9, and Y in position 11" hypothesis. Clearly the distinctive properties in 1-911-ANALOGY are responsible for the subjects' choices.

Although **T&G**'s model can discover distinctive features utilizing the size principle, it is limited to discovering the distinctive features of the base of the comparison (in T&G's framework, similarity is based on the intrinsically asymmetrical function of generalization, which depends only on the distinctive features of the base and not of the target). But for the subjects, the outcome of this problem depends on the distinctive features of the target (the opposite of what T&G predict). It seems unlikely, given the flexibility and sophistication of human thought, that all comparison processes will be bound by the asymmetrical properties of Bayesian inference. Further, if the model is extended to be able to perform bi-directional comparisons, how will it decide which of the computations to choose as the measure of similarity? Unless some principled way is specified, the model will be able to predict anything (and as such will explain nothing). It would appear that the model's predictions (asymmetrical comparison and the size principle) are not borne out by data. Rather, the hand-coded hypothesis space (a kind of a clairvoyant homunculus that can mysteriously assemble itself to fit any given occasion) carries most of the explanatory power.

Finally, we should evaluate any model not only on whether or not it can be falsified, but also, importantly, on its usefulness. How much does it add to our understanding of cognition? **T&G**'s model is only viable if we can somehow anticipate (and hand-code in) all the adjustments to the hypothesis space that will be required in any given situation (i.e., build in complete world knowledge). As such, the framework is either computationally unimplementable (if we can't build everything in) or psychologically uninformative (if we can).

A theory that applies equally well to all possible situations may apply poorly in each. This is especially true if generality requires us to disregard much of our hard-won understanding of the details of psychological processing. There is a vast literature documenting the complexity and diversity of representations and processes involved in similarity and categorization. The sheer variety of these psychological phenomena weighs heavily against any simple unitary account. Any such account can at best aspire to be a theory of spherical cows – elegant, but of little use in a world filled with cows that stubbornly insist on being cow-shaped.

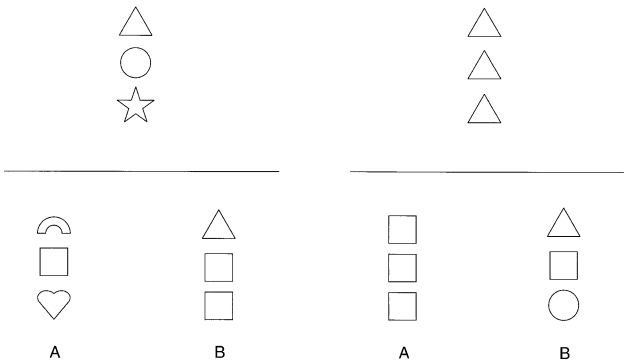



Figure 1 (Boroditsky & Ramscar). [This appears as Figure 6 in the article in this issue by Tenenbaum & Griffiths. The caption shown here is the caption written by T&G and appearing in T&G's article.] The relative weight of relations and primitive features depends on the size of the set of objects that they identify. Most observers choose B (the primitive feature match) as more similar to the top stimulus in the left panel, but choose A (the relational match) in the right panel, in part because the relation "all same shape" identifies a much smaller subset of objects than the relation "all different shapes."

ACKNOWLEDGMENTS

We would like to thank Alison Preston, Phillip Goff, Bradley Love, and Dedre Gentner for comments on earlier versions of this commentary.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

## Color constancy: A case for multiple levels and paradigms

Michael H. Brill

*Sarnoff Corporation, CN 5300, Princeton, NJ 08543-5300.*
**mbrill@sarnoff.com**

**Abstract:** Shepard claims that color constancy needs linear basis-function spectra, and infers the illuminant before removing its dependency. However, of the models of color constancy that have exact (and reasonable) spectral regimes, some do not need linear basis-function expansions of reflectance and illuminant spectra, some do not solve for the illuminant, and some estimate only partial object-reflectance information for single or multiple objects.
**[SHEPARD]**

To discuss color constancy, I must first (by assumption) exclude metamerism: If two reflectances match under illuminant "I" but not under illuminant "J," there is no visual transformation that can compensate the change from J to I. In **SHEPARD**'s world of more than three reflectance basis functions, my assumption would mean the eye is blind to all but a three-dimensional (3D) subspace of reflectances under all allowed illuminants.

Having said this, I think **SHEPARD** aptly describes a natural illuminant spectrum as "a terrestrial transformation of the invariant solar source." However, I do not agree with him that, to extract useful reflectance information from a scene, a visual system must find the illuminant transformation, invert it, and retrieve three reflectance-dependent quantities. Furthermore, even these tasks do not require (as he implies) that the illuminant and reflectance spectra are linear expansions of limited sets of basis functions.

Some investigators *define* color constancy using **SHEPARD**'s constraints, but this choice denies possibilities for strong invariance – under very general conditions. Furthermore, there is evidence that cognitive universals need not be represented one-to-one on perceptual space (one color for each reflectance): people report scene colors differently if asked "what is the color of the light?" as opposed to "what color is that surface?" (Arend & Reeves 1986). Hence we see illuminant biases in a scene, even though we also see illuminant-invariant attributes. To extrapolate, I think we might answer still differently if asked for particular aspects of a colored surface (such as chromaticness) or for color relationships among parts of an object.

***1. Older models outside Shepard's framework.*** Here are some examples of other color-constancy models that have been available for some time (see Brill & West 1986 for a review). A nonlinear model that first solves for the illuminant (Nikolaev 1985) assumes the illuminant and reflectance spectra are Gaussians in a monotonic function of wavelength. The parameters of the Gaussians are analogous to the basis-function coefficients of the linear models. A model that assesses only reflectance *relationships* (Brill & Hemmendinger 1985) makes no illuminant assumptions, but depends on the fact that the spectrum locus in chromaticity space is convex. In this model, the only invariant quantity is the right- or left-handed ordering of the chromaticities of three reflectances. Finally, there is a model that assesses reflectance relationships but does not assume the reflecting objects are coplanar (Petrov 1992). This model begins to come to terms with shading and shadows. However, like all the others discussed so far, it compares points in space, assuming that illuminants vary more smoothly in space than reflectances. This is a problem, for cast shadows and material boundaries have the same sizes and shapes (Arend 1994).

***2. A new model.*** The above problem can be avoided by posing an illuminant-invariant map based on the analysis of a single point in space. Let x be a monotonic function of visible wavelength. Suppose, at one point on the retina, the sensor values are R, G, B, and the sensors have peak x-values r, g, b. Define at this point the following function of R, G, B:

$$P = (g - b)log(R) + (b - r)log(G) + (r - g)\log(B). \quad (1)$$

Although difficult at first to believe, there are several alternative sets of spectral conditions under which $P$ depends only on reflectance and not on illuminant spectrum:

*Regime 1.* Equation 1 was originally applied to von Kries ratios rather than to R, G, B (Brill & West 1981). Later (Finlayson et al. 2000), it was applied to R, G, B. In both cases the following assumptions were used. Let the visual spectral sensitivities be equal-spread Gaussians in x, with standard deviation equal to 1 (chosen, without loss of generality, as the natural unit for x). Let the reflectance spectrum be Gaussian in x:

$$S(x) = a\, exp[-(x - p)^2/(2\, s^2)] \quad (2)$$

Finally, let the illuminant spectrum be exponential in x:

$$E(x) = c\, exp(f\, x). \quad (3)$$

Here, a, p, s, c, and f are coefficients.

Then, $P$ is invariant to illuminant change (change of c and f), but depends on the reflectance parameter *s (i.e., gives incomplete information about the reflectance):*

$$P = -0.5[(g - b)r^2 + (b - r)g^2 + (r - g)\, b^2]\, /(s^2 + 1). \quad (4)$$

*Regime 2.* Equation 1 is also illuminant-invariant for completely general reflectance spectra, provided the following conditions are satisfied (Finlayson et al. 2000; Marchant & Onyango 2000): The sensor spectral sensitivities are narrow-band, approaching delta functions in wavelength, and the illuminant spectrum have the following form:

$$E(x) = e\, F(x)\, exp(-fx). \quad (5)$$

[A noteworthy special case: for the Wien approximation to a black-body radiator (Wyszecki & Stiles 1982, p. 13), x = 1/wavelength, e and f depend on the black-body temperature T, and $F(x) = x^3$].

***3. Outlook.*** If we generalize **SHEPARD**'s definition of color constancy to include invariants that do not span the three dimensions of color, metamerism might yet be allowed in a color-constant system, so long as the additional freedom in the reflectance does not affect the values of the invariants. More generally, perceptual incompleteness of cognitive universals is compatible with vision models that make multiple incomplete representations of a scene (e.g., Lubin 1995). Such representations may be needed in our variegated world, allowing several hypotheses for visual truth to compete as reality unfolds.

## Colour perception may optimize biologically relevant surface discriminations – rather than type-I constancy

Nicola Bruno[a] and Stephen Westland[b]

[a]*Dipartimento di Psicologia, Università di Trieste, 34143, Trieste, Italy;* [b]*Colour and Imaging Institute, University of Derby, Derby, DEE2 3HL, United Kingdom.* **nicola.bruno@univ.trieste.itd**
**http://www.psico.univ.trieste.it/users/nick/**
**s.westland@colour.derby.ac.uk**
**http://www.colour.derby.ac.uk/colour/people/westland**

**Abstract:** Trichromacy may result from an adaptation to the regularities in terrestrial illumination. However, we suggest that a complete characterization of the challenges faced by colour perception must include changes in surface surround and illuminant changes due to inter-reflections between surfaces in cluttered scenes. Furthermore, our trichromatic system may have evolved to allow the detection of brownish-reddish edibles against greenish backgrounds.
**[SHEPARD]**

***Introduction.*** Human colour perception has evolved as a trichromatic system with specific receptoral sensitivities and post-receptoral transformations. **SHEPARD** is almost certainly right in propos-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ing that ecological forces have played a crucial role in shaping such a system. However, he may be wrong in characterizing these forces in terms of the inherent three-dimensionality of variations in the power spectra of natural illumination. Two arguments are raised against SHEPARD's view. First, a complete characterization of the challenges faced by colour perception must include not only illuminant changes, but also changes in surface surround, and illuminant changes due to inter-reflections between surfaces in cluttered environments. We claim that the sole ability to compensate for variations in illumination power spectra is probably inadequate to produce adaptive surface colours. Second, a number of recent results on the statistics of natural reflectance spectra and their relationship to human spectral sensitivities suggest that cone sensitivities optimize surface discriminations that were biologically important to our progenitors, most notably, those involving red-green discriminations. Under this view, the approximate colour constancy of the human visual system derives from the need to guarantee that such discriminations can be performed, rather than being a major evolutionary goal that required the internalization of the global statistics of reflectance and illumination variations.

***Two types of constancy in the light of mutual illumination.*** The environment in which our progenitors evolved was likely to be cluttered with natural formations of various kinds and was subjected to circadian variations in the spectral composition of daylight. To detect edible materials, such as fruit or roots, our species evolved the ability to use information in colour signals (Mollon 1989; Osorio & Vorobyev 1996) and the spatial relationships between colour signals (Foster & Nascimento 1994; Nascimento & Foster 1997). This ability amounts to solving three related challenges: achieving colour descriptors for surface materials despite changes in phases of illuminaton (Type-1 constancy); achieving constant descriptors despite changes in the surrounds (Type-2 constancy); and properly treating changes in intensity and spectral composition of the illumination due to mutual inter-reflections, shadowing, and transparency effects. The first challenge could conceivably be solved by exploiting statistical constraints on the variability of the phases of daylight (Shepard 1994). However, it is doubtful that the other two could. In fact, there is some consensus that solving the Type-2 constancy problem entails exploiting regularities in the distribution of surface reflectances, possibly using maxima in the distribution of colour signals (e.g., McCann 1992) or their variability (Brown & MacLeod 1997). In addition, there is a growing consensus that colour constancy will eventually require taking into account spatial structure (Schirillo 1999). In this respect, a standing problem for the field of colour vision is to connect a number of important facts that have emerged from the study of such effects of spatial structure in the perception of achromatic colours (Agostini & Galmonte 1999; Bruno et al. 1997; Cataliotti & Gilchrist 1995).

***How is the sampling of colour signals "optimal"?*** Since the pioneering contributions of Cohen (1964) and Maloney (1986), attempts at measuring the statistics of natural reflectance spectra have been performed in several laboratories (Parkkinen et al. 1989; Westland et al. 2000). Two crucial questions have been raised: how many basis functions are practically necessary to fully capture the variability of natural reflectances, and how does the abstract space defined by such bases relate to the coding of colour signals by the cones and the chromatically-opponent channels in the visual system? Answers to the first question have varied from three (Cohen 1964) to as many as twelve (Westland et al. 2000), depending on the intepretation one gives to the word "practically." Early answers to the second question (e.g., Buchsbaum & Gottschalk 1984) suggested that the first three basis functions are closely related to a luminance channel, red-green opponency, and yellow-blue opponency. Underlying this characterization of the mutuality of bases and opponent coding is the implicit assumption that chromatic coding is optimized to recover surface reflectance from the image intensity equation. However, whether this early answer is true in general is presently not clear. In a recent set of measurements (Castellarin 2000) on a large sample of natural reflectance spectra collected in Italy and the UK, we consistently found the first basis function to be an approximately increasing monotonic function of wavelength, which closely mirrors the sample average of our measured reflectances, not luminance. On the other hand, we find the second base to be highly similar to a red-green opponent signal; whereas the third and the fourth base show a much weaker relation to luminance signals and to yellow-blue opponent code. Our findings seem consistent with the notion that chromatic coding is optimized to capture a single dimension of variation in natural spectra, the red-green dimension, rather than fully reconstructing spectra reflectances. A similar proposal has been advanced by Nagle and Osorio (1993) using different statistical techniques and a different sample. The ability to perform most accurate discriminations along the red-green dimension may reflect pressures from the terrestrial environment of our progenitors, who hunted and gathered to detect brownish-reddish edibles against greenish backgrounds.

## Universal generalization and universal inter-item confusability

Nick Chater,[a] Paul M. B. Vitányi,[b] and Neil Stewart[c]

[a]*Institute for Applied Cognitive Science, Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom;* [b]*Centrum voor Wiskunde en Informatica, 1098 SJ, Amsterdam, The Netherlands;* [c]*Department of Psychology, University of Warwick, Coventry, CV4 7AL, United Kingdom.* **nick.chater@warwick.ac.uk paulv@cwi.nl neil.stewart@warwick.ac.uk http://www.warwick.ac.uk/fac/sci/Psychology/staff/academic.html#NC/ http://www.cwi.nl/~paulv/ http://www.warwick.ac.uk/fac/sci/Psychology/staff/postgraduates.html#NS/**

**Abstract:** We argue that confusability between items should be distinguished from generalization between items. Shepard's data concern confusability, but the theories proposed by Shepard and by Tenenbaum & Griffiths concern generalization, indicating a gap between theory and data. We consider the empirical and theoretical work involved in bridging this gap. **[SHEPARD; TENENBAUM & GRIFFITHS]**

**SHEPARD** shows a robust psychological law that relates the distance between a pair of items in psychological space and the probability that they will be confused with each other. Specifically, the probability of confusion is a negative exponential function of the distance between the pair of items. In experimental contexts, items are assumed to be mentally represented as points in a multidimensional Euclidean space, and confusability is assumed to be determined according to the distance between items in that underlying mental space. The array of data that **SHEPARD** amasses for the universal law has impressive range and scope.

Although intended to have broader application, the law is primarily associated with a specific experimental paradigm – the identification paradigm. In this paradigm, human or animal agents are repeatedly presented with stimuli concerning a (typically small) number of items. We denote the items themselves as a, b, . . . , corresponding stimuli as S(A), S(B), . . . , and the corresponding responses as R(A), R(B), . . . . People have to learn to associate a specific, and distinct, response with each item – a response that can be viewed as "identifying" the item concerned.

How does a law concerning confusability in the identification paradigm relate to the question of generalization? We suggest that there is no direct relationship. Generalization from item A to item B in the sense discussed by **SHEPARD**, involves deciding that an item b has property f, because item a has property f. This is an inductive inference: f(A), therefore f(B). By contrast, confusing item A with B means misidentifying item A as being item B. Generalization typically does not involve any such misidentification: on learning that a person has a spleen, I may suspect that a goldfish has a spleen – but there is no need to misidentify or mix up people and goldfish.

These observations suggest that there may be a gap between



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

**SHEPARD**'s theoretical analysis, which considers the question of generalization, and his empirical data-base, which concerns confusability. This points up two distinct research projects, attempting to reconnect theory and data.

The first project attempts to connect theory to data. This requires gathering empirical data concerning generalization, to see to what extent generalization does have the negative exponential form predicted from **SHEPARD**'s theoretical analysis. This project is, to a limited degree, taken up in **TENENBAUM & GRIFFITHS**' empirical studies of generalization from single and multiple instances. These preliminary results suggest that the generalization function appears to be concave, which also fits with their Bayesian theoretical analysis. Whether the data have an exponential form, and whether there is a universal pattern of data across many different classes of stimuli, must await further empirical work. But some of our own results have suggested that generalization may be surprisingly variable, both between individuals and across trials, even with remarkably simple stimuli.

Stewart and Chater (submitted) investigated generalization to novel stimuli intermediate between two categories that differ in variability. The effect of the variability of the categories differed greatly between participants – some participants classified intermediate stimuli into the more similar, less variable category; others classified the intermediate stimuli into the less similar, more variable category. Further, altering the variability of the training categories had large effects on individual participants' generalization. When the difference in variability between the two categories was increased, some people increased generalization to the more variable category, and some increased generalization to the less variable category. Extant exemplar (e.g., Nosofsky 1986) and parametric/distributional (e.g., Ashby & Townsend 1986) models of generalization in categorization cannot predict the large variation between participants. This individual variation in performance suggests that there may be no single law governing human generalization, and therefore, that performance may not fit into **SHEPARD**'s theoretical analysis, although it is too early to draw firm conclusions on this issue.

The second project arises from the apparent gap between theoretical analysis and empirical data in **SHEPARD**'s program concerns connecting data to theory. SHEPARD has provided a strong evidence that confusability is an inverse exponential function of distance in an internal multidimensional space. How can this result be explained theoretically? The rest of this commentary develops a possible approach.

To begin with, we note that the view of psychological distance as Euclidean distance in an internal multidimensional space may be too restrictive to be applicable to many aspects of cognition. It is typically assumed that the cognitive representation formed of a visually presented object, a sentence or a story, will involve structured representations. Structured representations can describe an object not just as a set of features, or as a set of numerical values along various dimensions, but in terms of parts and their interrelations, and properties that attach to those parts. For example, in describing a bird, it is important to specify not just the presence of a beak, eyes, claws, and feathers, but the way in which they are spatially and functionally related to each other. Equally, it is important to be able to specify that the beak is yellow, the claws orange, and the feathers white – to tie attributes to specific parts of an object. Thus, describing a bird, a line of Shakespeare, or the plot of Hamlet as a point in a Euclidean multidimensional space appears to require using too weak a system of representation. This line of argument raises the possibility that the Universal Law may be restricted in scope to stimuli which are sufficiently simple to have a simple multidimensional representation – perhaps those that have no psychologically salient part-whole structure. We shall argue, however, that the Universal Law is applicable quite generally, since all these aspects are taken into account by the algorithmic information theory approach. This leads to a more generalized form of the Universal Law.

In particular, we measure the distance between arbitrary representations (whether representations of points in space, of scripts, sentences, or whatever), by the complexity of the process of "distorting" each representation to the other. Specifically, the distance between two representations, A and B, is defined to be the *sum* of the lengths of the shortest computer program that maps from A to B *and* the length of the shortest computer program that maps from B to A. This is known as sum-distance (Li & Vitányi 1997). Sum-distance measure is attractive not only because it has some theoretical and empirical support as a measure of similarity (Chater & Hahn 1997; Hahn et al., submitted), but also because it connects with the theoretical notion of information distance, developed in the mathematical theory of Kolmogorov complexity (Li & Vitányi 1997). (See Chater 1999, for an informal introduction in the context of psychology.) The intuition behind this definition is that similar representations can be "distorted" into each other by simple processes, whereas highly dissimilar representations can only be distorted into each other by complex processes; the complexity of a process is then measured in terms of the shortest computer program that codes for that process.

**SHEPARD** uses a specific function, G(A, B), as a measure of the confusability between two items. It turns out that – using only the assumption that the mapping between the input stimuli and the identification responses is computable – it can be shown that G(A, B) is proportional to the negative exponential of the sum-distance between A and B. That is, if distance is measured in terms of the complexity of the mapping between the representations A and B, then SHEPARD's universal law, when applied to confusability, follows automatically (Chater & Vitányi, submitted).

We have suggested that this result is attractive, because it applies in such a general setting – it does not presuppose that items correspond to points in an internal multidimensional psychological space. This observation suggests a further line of empirical research: to determine whether the Universal Law does indeed hold in these more general circumstances.

## Generalization and Tinbergen's four whys

Ken Cheng

*Department of Psychology, Macquarie University, Sydney NSW, 2109, Australia.* **kcheng@axon.bhs.mq.edu.au**
**http://www.axon.bhs.mq.edu.au/kcheng/homepageofKen.html/**

**Abstract:** Shepard's exponential law provides a functional explanation of generalization. The account complements the more common mechanistic models. The elegant and powerful analyses answer one of Tinbergen's (1963) four whys of behavior: a benefit conferred on the animal by generalizing in this way. A complete account might address evolutionary and developmental questions in addition to mechanistic and functional ones. **[SHEPARD]**

In the classic paper "On aims and methods in Ethology," Tinbergen (1963) identified four types of "why" questions to be addressed about any behavior. Mechanistic explanations concern the immediate conditions for a behavior, from stimulus conditions to brain structures. Developmental questions ask about the ontogenetic history of a behavior. The field of psychology addresses mostly these proximate questions of mechanism and development, with more on mechanism than on development. Far less frequently tackled are the ultimate questions of function (what Tinbergen called "survival value") and evolution. Functional questions address the benefits conferred, at present, by a behavior, while evolutionary questions address the evolutionary history of a behavior. To fully understand a phenomenon in learning, perception, or cognition, answers to all four whys are needed.

**SHEPARD**'s article reprinted here takes the road less travelled and answers functional questions about learning, cognition, and perception. The aim is to look for abstract universal principles that animals "should" honor because the world they live in possesses


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

certain invariant properties. To put it bluntly, behaving in accord with these invariants should add survival value. The invariants are abstract and deep, and digging them out is hard work to my mind. The functional universals thus unearthed add a whole new dimension to understanding the phenomena, a dimension often missing in psychology.

In the rest of my commentary, I will limit consideration to the topic of generalization, the topic with which I am most familiar. Mechanistically, various models of spreading activation, going back to Shepard (1958), can produce generalization gradients (e.g., Cheng et al. 1997; Reid & Staddon 1998). Others take a network approach (e.g., Ghirlanda & Enquist 1998; 1999; Gluck 1991; Saksida 1999). Choosing amongst them remains difficult, but we have plenty of recent thinking on the topic.

On the functional question, **SHEPARD**'s is the only account of generalization to date. It offers far more than speculation about the possible advantages of generalizing. The analysis tackles the form of gradients. It is SHEPARD's style not to contrast how animals might differ in generalization, but to find universals. In the face of seeming diversity, SHEPARD tells us where and how to look for universality. The functional analysis came up with elegant reasons why animals should follow the exponential law, and the conditions and idealizations required for finding it. It is thus powerful in offering not only reasons for generalizing, but some deep insights into the way it should happen. The law has found supporting evidence in humans and pigeons (Shepard 1987b), and recently in honeybees (Cheng 2000).

The exponential law gives us a universal for generalization, and tells us why it benefits animals today. We may further ask how animals evolved to generalize in this way. Given that the law is found in diverse animals, convergent evolution is suggested. But it is hard to add much. The evolutionary question is difficult to answer for the lack of behavioral records. Generalization gradients are not imprinted on rocks. An answer will likely require a far broader and deeper comparative study of learning.

How do animals "come up with" **SHEPARD**'s law in the course of their lives? One possibility is that $y = e^{-kx}$ is wired in the brain. To be more precise, the exponential law might be mostly a matter of maturation. The worker bee hatching out of her cell immediately starts generalizing in accord with SHEPARD's law, in each and every task that she undertakes in her life. Thus, the initial state for generalization is an exponential gradient. Experience fills in the scaling parameter left free in the equation (k). But the initial state might be a broader class of functions, and experience might be necessary to narrow the gradients down to the exponential shape. Evolution does not have to wire in the equation, so to speak. It just has to ensure that the animal would arrive at the equation in the course of a typical life. As an example, consider the migration of the indigo bunting in its first autumn of life. The bunting is known to fly south by using stars at night. It has to know which way is south by the pattern of the night sky in the northern hemisphere. Is a map of stars etched into the brains of indigo buntings to guide them south? Classic work by Emlen (1975) shows that this is not so. Using a planetarium to manipulate night sky conditions, Emlen showed that some exposure to the night sky before the time of migration is necessary for oriented navigation. It turns out that the birds, from viewing the night sky, extract the fixed point of rotation of the stars, and fly in a direction opposite to that. Data on the issue of the development of generalization are lacking. The topic is perhaps best examined in an insect forager such as the worker honeybee, whose life cycle is short, whose learning is quick, and whose experiences can be controlled to a great extent.

In sum, I very much welcome **SHEPARD**'s functional approach to cognition. In addition to mechanistic questions, all cognitive science, comparative cognition included, should take aboard functional questions into research. We should also address developmental and evolutionary questions.

## Which colour space(s) is Shepard talking about?

Lieven Decock and Jaap van Brakel

*Department of Philosophy, University of Leuven, 3000 Leuven, Belgium.*
**{Lieven.Decock; Jaap.vanBrakel}@hiw.kuleuven.ac.be**

**Abstract:** Contra Shepard we argue, first, that his presentation of a three-dimensional representational (psychological or phenomenal) colour space is at odds with many results in colour science, and, second, that there is insufficient evidence for Shepard's stronger claim that the three-dimensionality of colour perception has resulted from natural selection, moulded by the particulars of the solar spectrum and its variations.
**[SHEPARD]**

According to **SHEPARD** the colour appearances of surfaces correspond to relatively fixed points in a three-dimensional *colour space* (his emphasis). However, the distinction between phenomenal, perceptual, psychological, or internalised representational colour spaces and the various technological or (psycho)physical colour spaces is blurred. He takes as self-evident that these colour spaces are isomorphic. Examples of psychophysical colour spaces are the CIE chromaticity diagram, a wavelength mixture space (Clark 1993, p. 37), and a retinex colour space (Land 1986, p. 12). Such colour spaces are characterised by means of a limited number of parameters that can be computed on the basis of precise measurements by means of spectrometers and underlying physical theory. SHEPARD's internalised representational colour space, based on the traditional account of a perceptual colour space, is not likely to be one of these "physical" spaces. Furthermore SHEPARD's representational space-time space may have been constructed on analogy with the internal colour space. Here, however, we will focus on the three dimensionality of his colour space.

First, it is by no means clear how to give a good operational characterisation of the three dimensions of **SHEPARD**'s representational colour space in terms of lightness, hue, and saturation. The difficulty is very apparent in the ubiquitous ambiguity surrounding a lightness- or brightness-axis. The lightness-axis is, primarily, a black-white axis, based on contrast experiences; the brightness axis is based primarily on the luminosity of colour patches. Problems arise with the psychological difference between black/white, dark/light, and dull/bright. Furthermore, there are interdependencies between hue, brightness, and saturation, however defined, and the three of them fail to cover all aspects of "colour" appearance (for references see Saunders & van Brakel 1997, p. 175f).

Second, it is not obvious how to characterise the dimensions of the representational colour space unambiguously. Traditionally, there are three dimensions, but this rests on rather vague introspective intuitions. Multi-dimensional scaling (MDS) techniques allow the ordering of colour comparisons in a spatial structure. However, these techniques yield a variety of results and are difficult to interpret. Moreover, how to choose samples that do not prejudge the outcome? Further, even if MDS techniques yield three dimensions, there is nothing to tell you how to define the axes and measure distances. Finally, it has been claimed that four, six, or seven dimensions are needed to adequately represent human colour vision (Chang & Carroll 1980; Sokolov 1997).

Third, **SHEPARD** presents the colour space as "approximating the idealised spherical solid." Although his characterisation is hedged, it still suppresses the many proposed "forms" of colour space. It has been presented as an infinite cylinder – the hue-saturation circle remaining but the brightness (dark-dazzling) axis being infinite (Thompson 1995, p. 47). It can also be presented as a cylinder with a finite lightness- or brightness-axis. Sivik's Natural Colour System is based on a double cone (Sivik 1997). All these are "neat" geometrical shapes. Moreover, empirical evidence pulls in the direction of less well-behaved spaces. The well-known Munsell colour space has a bulge in the purple area. There has been talk of "a Riemannian space with global cylindrical co-ordinates" or "a power



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

metric of Minkowski" (Indow 1988). After 30 years of work an OSA committee concluded that it is impossible to construct a (regular) rhombohedral lattice with uniform colour distances without curvature (Man & MacAdam 1989; Nickerson 1981).

We conclude, therefore, that there is too little information about the precise form or dimension of the perceptual colour space to allow well-defined mathematical transformations to be carried out. In view of the plurality of proposals, the absence of consensus over the global structure, and the lack of operational procedures to define axes or to measure in the space, one can even doubt the existence of a genuine representational colour space (Decock 2001; Mausfeld et al. 1992).

Moreover, even if the internalised colour space is simply equated with a wavelength mixture space, and trichromacy is construed as the fact that there are three receptors in the human eye, there remain several problems with **SHEPARD**'s assumption of universal three-dimensionality. Human trichromacy is related to the number of cones in the eyes, and SHEPARD mentions a suggestion of Maloney and Wandell (1986) that, in order to have more degrees of freedom, one would need more types of photoreceptors. However, trichromacy is much less universal than SHEPARD assumes. Putting aside worries about the difficulty of interpreting studies in comparative "colour" vision (Jacobs 1992), pigeons, turtles, some kinds of fish (e.g., the goldfish), and jumping spiders are tetrachromatic; some birds may be pentachromatic. In contrast, most primates are dichromatic, although the spider monkey is tetrachromatic. Among humans too, there is diversity. Dichromacy is often a result of genetics and not of some defect, and it has been suggested that some human females are intrinsically tetrachromatic (Jordan & Mollon 1993). There is more reason to believe that the evolution to the "dimension" of colour vision is determined more by the ecology of the particular animal than by the universal planetary atmosphere (Nuboer 1986; Thompson 1995, pp. 190–95).

It is not merely a matter of disagreeing about plausible speculations. There are also more technical problems with **SHEPARD**'s proposals. There is a strong tendency to overinterpret mathematical and physical approximations. For example, results by Judd et al. (1964), later corroborated by others, establish that daylight can fairly well be given by three functions. A sample of 622 daylight spectra was taken, and subjected to an eigenvector analysis. Most of the variation was given by three functions. However, the number obtained is relative to the degree of accuracy wanted: an analysis of the residual information would have led to a fourth dimension; two functions might already give a "reasonable" degree of accuracy. Furthermore, in order to calculate these functions important restrictions were made. For example, the 622 chosen spectra were smoothened and a decision too was made on the bandwidth by which daylight was measured. That a more fine-grained spectrum could be represented to the same degree of accuracy by means of three functions is unlikely. In fact, several studies have been reported that indicate that not all of the variation can be described by three functions (Wyszecki & Stiles 2000, p. 11, mention nine studies).

A similar overinterpretation of mathematical approximation techniques is to be found in the linearity assumption. Grassman's laws support a linearity assumption. The mathematical theory is well elaborated in Krantz (1989). But it is difficult to believe that this assumption is more than a good approximation. The difficulties in finding a global uniform colour space are an immediate indication of its limitations. In colour science reports, non-linearities are ubiquitous. For example, the spectral absorbance curves of the cones are neither linear nor stable (because of bleaching effects). Contrast effects subsequently distort the processing of this information. Further along the visual system, neurons would execute the linear mathematical operations. However, they do not behave like analogous electrical devices or like digitalised computers; instead, their firing above threshold excitation resembles a nonlinear Heaviside-function. At best, the linear transformations **SHEPARD** (following Maloney) describes are a reasonable approximation of what is going on in a black box.

The three functions that describe the daylight spectral variations are overinterpreted too. They are associated with different atmospheric conditions. But this is only a crude comparison (Wyszecki & Stiles 2000, p. 11). The three atmospheric conditions considered are not the most typical. It is not clear, for example, why the daylight from the solar disk alone, and of the solar disk at low altitudes plus sky, should, in general, be the same. It is also not clear that the situations sketched in Wyszecki and Stiles are the same as **SHEPARD**'s. The curves were computed from a "random" set of 622 spectral samples, not on the basis of physical (atmospheric) information.

## Universal Bayesian inference?

David Dowe[a] and Graham Oppy[b]

[a]*Department of Computer Science, Monash University, Clayton, Vic 3800, Australia;* [b]*Department of Philosophy, Monash University, Clayton, Vic 3800, Australia.* **David.Dowe@infotech.monash.edu.au**
**Graham.Oppy@arts.monash.edu.au**
**http://www.cs.monash.edu.au/~dld/**

**Abstract:** We criticise Shepard's notions of "invariance" and "universality," and the incorporation of Shepard's work on inference into the general framework of his paper. We then criticise Tenenbaum and Griffiths' account of Shepard (1987b), including the attributed likelihood function, and the assumption of "weak sampling." Finally, we endorse Barlow's suggestion that minimum message length (MML) theory has useful things to say about the Bayesian inference problems discussed by Shepard and Tenenbaum and Griffiths.

**[BARLOW; SHEPARD; TENENBAUM & GRIFFITHS]**

**SHEPARD** (1994; target article) claims that it is a general fact about the world that objects which are of the same basic kind generally form a connected local region in the space of possible objects. *Prima facie,* at least, this is not a "fact about the world" at all; rather, it is an analytic or *a priori* truth which connects together the notions of "basic kind" and "connected local region in the space of possible objects." Moreover, even if this were a general fact about the world, it seems implausible to suppose that it would have an important role to play in the explanation of the "universality" of Bayesian inference: the efficacy of Bayesian inference does not depend upon the *kinds* of things which there are. The "universality, invariance and mathematical elegance" which SHEPARD finds for this theory of inference seems different in kind from that which he finds for his theory of perceived colour – and perhaps also for his theory of perceived motion – even though it might be accommodated in an appropriate evolutionary theory (on the grounds that Bayesian inferrers will be advantaged in any possible world in which mobile perceivers can evolve).

**SHEPARD**'s use of the term "invariance" – in characterising the general aim of this theory – puzzles us. He claims that the General Theory of Relativity provides a model for the kind of psychological theory which he wants. However, it seems to us that there are several confusions in this part of his discussion. Most importantly, while it is true that Einstein cast General Relativity in generally covariant form, it is perfectly possible to cast Newtonian mechanics in generally covariant form: this just amounts to the observation that it is possible to give a coordinate-free formulation of these theories. In our view, the most natural notion of invariance – or invariance group – for physical theories is that developed by Anderson (1967), which turns on questions about the "absolute objects" postulated by these theories: the advance which is marked by General Relativity is that it contains no absolute objects. Furthermore, as we understand it, it was never the case that Newton's laws were restricted to inertial frames moving at slow speeds relative to the speed of light; rather, it turns out that Newton's "laws" do not fit the data for objects moving at speeds which are not slow relative to the speed of light. We confess that we have no idea what SHEPARD has in mind when he says that he is aiming for "invariant" psychological principles. (We have similar concerns



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

about his use of the term "universal," and its relations to "invariance.")

**TENENBAUM & GRIFFITHS** begin with "Shepard's problem of generalisation from a single positive instance." They propose a modification – to what they take to be **SHEPARD**'s approach – which makes a significant difference when we move on to consider generalisation from multiple positive instances. In their view (sect. 2.3), Shepard (1987b) argued for the default assumption that the positive instance was sampled uniformly from the full range of cases ("weak sampling"); **TENENBAUM & GRIFFITHS** claim that in many cases, it is better to suppose that the positive instance is sampled uniformly from the consequential region, that is, from amongst the positive cases ("strong sampling"). Moreover, TENENBAUM & GRIFFITHS suppose that this difference shows up as a difference in the likelihood functions for the two approaches. They favour the likelihood function which takes the value $1/|h|$ when $x \epsilon h$ and is otherwise 0; and they attribute to SHEPARD the likelihood function which takes the value 1 when $x \in h$, and is otherwise 0. We think that something has gone wrong here: the likelihood function attributed to SHEPARD has not been normalised, and the fact that it attributes probability 0 to data outside the consequential region contradicts the independence in their informal definition of "weak sampling." Moreover, this function makes no sense as a value for the conditional probability of observing x given that h is the true consequential region: under the assumption that the sampling is uniform from the entire range of cases, that value would have to be $1/|C|$, where C is the measure of the entire space. Perhaps TENENBAUM & GRIFFITHS have here conflated the likelihood function with the probability that the sample lies in the consequential region (which is indeed 1 if the sample is taken from the consequential region).

There may be reason for caution in attributing "weak sampling" to **SHEPARD**. It seems that SHEPARD does make this assumption (1987b, p. 1321) – though we didn't find SHEPARD's text entirely clear, and we have some sympathy for **TENENBAUM & GRIFFITHS'** interpretation of it – but only in the context of his "theoretical justification" for the choice of a probability density function for the *size* of the consequential region. What he says is that, *in the absence of any information to the contrary,* an individual might best assume that nature selects the consequential region and the first stimulus independently. However, in the cases which TENENBAUM & GRIFFITHS consider, it is plausible that SHEPARD would deny that there is no further information to the contrary – the baby robin has reason to suppose that mother will have sampled from the consequential region; and the doctor knows that the patient has been sampled from the consequential region – so it is not clear that SHEPARD cannot get the same answers as TENENBAUM & GRIFFITHS in these cases. Moreover, TENENBAUM & GRIFFITHS discuss the choice of a probability density function for the *identity* of the consequential region; so they are not discussing exactly the same question which was taken up by SHEPARD. [It is curious to us that TENENBAUM & GRIFFITHS use Erlang priors in their examples: after all, in their view, SHEPARD's "rational justification" for this choice of prior *relies upon* the assumption of weak sampling. From their point of view, it requires work to show that these priors admit of "rational justification" under the assumption of strong sampling; and we would add that there are mathematically convenient priors which do not admit of rational justification (see Wallace & Dowe 1999b, pp. 334–35). On an unrelated point, it is also curious to us that TENENBAUM & GRIFFITHS claim that *generality* is more primitive than *similarity* (sect. 4.1) – but then go on to say that judgments about similarity are required in order to provide reasonable constraints on generalisation (sect. 5). We doubt that there is any neat separation of problems to be made here: see Wallace and Boulton (1973) for an example of the simultaneous analysis of both problems in a Bayesian machine-learning context.

Perhaps **TENENBAUM & GRIFFITHS** might reply that they have given an extension of **SHEPARD**'s work to an area which SHEPARD himself had not considered: given cases in which the natural assumption is strong sampling, what kinds of probability distributions would we expect creatures to have? However, as TENENBAUM & GRIFFITHS note, there are many other kinds of cases in which various other natural assumptions may or may not be made. And, in the end – though TENENBAUM & GRIFFITHS do not put it this way – perhaps all that they commit themselves to is the claim that the most natural *general* extension of SHEPARD's work supports nothing more than the hypothesis that Bayesian inference is universal. Yet, if that is right, then it is hard to see that we are getting "universal, invariant general principles" which reflect the internalisation of "pervasive and enduring facts about the world." Perhaps we might hold that we are getting "universal, invariant, general principles": though even that seems a bit of a stretch. After all, it is one question what are the optimal inferences to be made from given data; it is quite another question how close we should expect evolved creatures to come to these optimal inferences in particular cases. There seems to be little reason to suppose that evolved creatures will be perfect Bayesians across the board; indeed, we know from countless experiments on people that we are very far from being perfect Bayesian reasoners ourselves. Unless we are prepared to hold that you don't count as a "perceptually advanced mobile organism" unless you are a reliable Bayesian information processor, we see little reason to suppose that there is a *universal* law of the kind which SHEPARD proposes (even allowing a restriction to cases in which weak sampling can be presupposed). Opponents of Bayesian theories of inference will no doubt also have many reasons for wishing to raise objections here.

**TENENBAUM & GRIFFITHS** appear to claim that their theory "uniquely" extends **SHEPARD**'s work in a Bayesian framework (sect. 3.3). However, one can use a Bayesian approach to infer the consequential region itself: see, for example, Wallace and Dowe (1999b, pp. 332–34) for analysis of a related continuous case in which the size of the consequential region is known.

We agree with the implicit suggestion of **BARLOW** that MML theory might offer the best Bayesian theory of inference – see Wallace and Boulton (1968), Wallace and Freeman (1987); and see Wallace and Dowe (1999a) (and the rest of this 1999 special issue of the *Computer Journal*) for a discussion of the contrast between minimum message length, minimum description length, and the work of Solomonoff (1964a; 1964b) and others. To the extent that we expect evolution to optimise, we should expect to find MML inferences in nature. No doubt we often do. MML can be made to fit the empirical data described by Shepard (1987b). And it can be made to yield the generalisations offered by **TENENBAUM & GRIFFITHS**. And it has a high degree of mathematical elegance, and so on. But we do not think that it would help the general thesis which **SHEPARD** defends to espouse MML – even though the suggestion could be quite congenial to him. That Bayesian MML inference is evolutionarily advantageous is true, but true quite independently of facts about the *kinds* of objects that exist in the world.

## External regularities and adaptive signal exchanges in the brain

Birgitta Dresp

*LSBMC, IMFS, UMR 7507, Université Louis Pasteur – Centre National de la Recherche Scientifique, Strasbourg, France.*
**birgitta.dresp@ensps.u-strasbg.fr**

**Abstract:** Shepard's concept of internalization does not suggest mechanisms which help to understand how the brain adapts to changes, how representations of a steadily changing environment are updated or, in short, how brain learning continues throughout life. Neural mechanisms, as suggested by Barlow, may prove a more powerful alternative. Brain theories such as Adaptive Resonance Theory (ART) propose mechanisms to explain how representational activities may be linked in space and time. Some predictions of ART are confirmed by psychophysical and neurophysiological data.
**[BARLOW; SHEPARD]**



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

The target articles in this special *BBS* issue discuss the essential question of how the brain generates representations of the outside world. There is general agreement that the probabilistic processing of regularities in the environment is important for the adaptation and survival of both man and animal. To account for the emergence of representations of physical regularities, **SHEPARD** suggests that the brain learns about the laws that govern the relations between objects and their occurrence in space and time, and that the internalization of these relations, or laws, leads to stable perceptions and reliable cognitive representations of the outside world. Observations such as the apparent motion of an object when it is first presented at one position, and then at another position in space, and the possibility to relate this perceptual phenomenon to the laws of kinematic geometry are used to back up the theory. The concept of internalization implies that the brain stores multiple copies of objects and events and all their possible relations in space and in time. As **BARLOW** and other authors here point out, statistical regularities are important for learning and memory. Learning continues throughout life, and learnt representations have to be continuously updated because we live in a world that keeps on changing. How would internalization account for the fact that the brain has the capacity to generate stable representations of regularities and events, but at the same time appears able to change or update these representations whenever it becomes necessary?

Adaptive Resonance Theory (ART; e.g., Grossberg 1999) suggests how a massively parallel, distributed neural network structure such as that which constitutes the brain may generate signal exchanges that produce representations of spatio-temporal regularities and adapt to significant changes rapidly. In such a theoretical framework, top-down memory representations of perceptual events continuously interact with ongoing bottom-up input to detect and generate representations of spatio-temporal coincidence. Repeated spatial or temporal coincidences between top-down expectation signals and bottom-up input signals reinforce the relative weights of signal exchanges in a given neural circuit. Such signal exchanges would indeed be able to generate stable representations of regularities via reinforcement, but they would also have the capacity to adapt rapidly to sudden changes and new situations.

The contrast detection of a visual target line presented with and without a structural context represents a good example of an experimental finding suggesting that such signal exchanges do occur in the brain. When a contrast target is presented alone on a screen, its detection threshold does not improve significantly with the progression of the trial blocks. However, when the same target is embedded in a context of co-linear lines, its detection threshold is found to decrease considerably with the progression of the trial blocks (Dresp 1993; 1999). This effect of co-linearity on the detection of local visual signals may signify that the extraction of a spatial regularity such as co-linearity is based on signal exchanges between orientation-selective neurons detecting spatial and temporal coincidence along a given axis of alignment in visual patterns, as ART would suggest. Co-linearity, or alignment, is a major grouping factor in Gestalt theory and an important regularity in visual scenes because co-linear fragments are likely to belong to the same visual object. It has been suggested that co-linearity is represented in the brain by neural "association fields" (Field et al. 1993). Neurophysiological data demonstrate that the firing rate of a V1 cortical neuron increases when a co-linear line is presented together with the line that probes the receptive field of the neuron under investigation (Kapadia et al. 1995). This increase in firing rate of a V1 neuron caused by a contextual stimulus presented outside its classic receptive field (Gilbert & Wiesel 1990) is indeed correlated with a decrease in the psychophysical detection threshold of the line probing that field (Kapadia et al. 1995). Such a context-related detection facilitation diminishes in a nonlinear manner as the distance between target line and co-linear context line increases, and the function describing these changes in thresholds is predicted by artificial neural network structures simulating lateral connectivity between neurons (Fischer et al. 2000).

As **BARLOW** points out, the way perceptual events are represented in the brain is far from being settled. He attracts attention to the question of how cortical neurons exploit the regular properties of external objects, and evokes possible neural mechanisms which link perceptual events in space and time. The idea that the brain may use such mechanisms to generate representations of regularity goes beyond **SHEPARD**'s concept of internalization and encourages us to look further in psychophysics, neurophysiology, and computational theory.

## Neural spaces: A general framework for the understanding of cognition?

Shimon Edelman

*Department of Psychology, Cornell University, Ithaca, NY 14853-7601.*
**se37@cornell.edu** **http://www.kybele.psych.cornell.edu/~edelman/**

**Abstract:** A view is put forward, according to which various aspects of the structure of the world as internalized by the brain take the form of "neural spaces," a concrete counterpart for Shepard's "abstract" ones. Neural spaces may help us understand better both the representational substrate of cognition and the processes that operate on it.
**[SHEPARD]**

**SHEPARD**'s meta-theory of representation, illustrated in the target article by three examples (object motion, color constancy, and stimulus generalization), can be given the following general formulation: the existence of an invariant law of representation in a given domain is predicated on the possibility of finding an "abstract space" appropriate for its formulation. The generality of this meta-theory stems from the observation that any sufficiently well-understood physical domain will have a quantitative description space associated with it. In the account of perceived motion, this is the constraint manifold in what is called the configuration space in mechanics (and robotics). In color vision, it is the low-dimensional linear space that can be related through principal component analysis to the characteristics of natural illumination and surface reflectances. In stimulus learning and generalization, it is the "probabilistic landscape" with respect to which Marr (1970) formulated his Fundamental Hypothesis[1] and over which SHEPARD's (1987b) "consequential regions" are defined.

A central thesis of the target article is that evolutionary pressure can cause certain physical characteristics of the world to become internalized by the representational system. I propose that the internalized structure takes the form of *neural spaces*, whose topology and, to some extent, metrics, reflect the layout of the represented "abstract spaces."[2]

The utility of geometric formalisms in theorizing about neural representation stems from the straightforward interpretation of patterns of activities defined over ensembles of neurons as points in a multi-dimensional space (Churchland & Sejnowski 1992; Gallistel 1990; Mumford 1994). Four of the issues stemming from the neural space (NS) approach to representation that I raise here are: (1) its viability in the light of experimental data; (2) the explanatory benefits, if any, that it confers on a theory of the brain that adopts it; (3) the operational conclusions from the adoption of the NS theoretical stance; and (4) the main theoretical and experimental challenges it faces.

***Viability.*** The relatively few psychophysical studies designed specifically to investigate the plausibility of attributing geometric structure to neural representation spaces did yield supporting evidence. For example, the parametric structure built into a set of visual stimuli can be retrieved (using multi-dimensional scaling) from the perceived similarity relationships among them (Cortese & Dyre 1996; Cutzu & Edelman 1996; Shepard & Cermak 1973). Psychophysical evidence by itself cannot, however, be brought to bear on the neurobiological reality of a neural space. To determine whether or not the geometry of a postulated neural space is


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

causally linked to behavior, one needs to examine the neural activities directly. Unfortunately, neurophysiological equivalents of the psychophysical data just mentioned are very scarce. The best known direct functional interpretation of neuronal ensemble response was given in connection with mental rotation in motor control (Georgopoulos et al. 1988). More recently, an fMRI study of visual object representation that used multidimensional scaling to visualize the layout of the voxel activation space yielded a low-dimensional map that could be interpreted in terms of similarities among the stimuli (Edelman et al. 1999).

***Potential benefits.*** The representation space metaphor has been invoked as an explanatory device in many different areas of cognition, from visual categorization (Edelman 1999) to semantics (Landauer & Dumais 1997). In vision, this move has been used, traditionally, to ground similarity and generalization. When a transduction mechanism connecting the neural space to the external world is specified, the geometric metaphor also provides a framework for the treatment of veridicality of representations (Edelman 1999), in a manner compatible with **SHEPARD**'s idea of second-order isomorphism between representations and their referents (Shepard & Chipman 1970). Conceptual spaces seem to offer a promising unified framework for the understanding of other aspects of cognition as well (Gärdenfors 2000).

***Operational conclusions.*** Adopting the NS idea as a working hypothesis leads to some unorthodox and potentially fruitful approaches to familiar issues in cognition. One of these issues, raised in the target article, is what branch of mathematics will emerge as the most relevant to the understanding of cognition in the near future. **SHEPARD** mentions in this context group theory; the work of **TENENBAUM & GRIFFITHS** suggests that Bayesian methods will be useful. If the spatial hypothesis is viable, cognitive scientists may also have to take up Riemannian and algebraic geometry. Another issue to consider is the basic nature of the information processing in the brain. Assuming that the representations harbored by the brain are intrinsically space-like, the model of computation best suited for the understanding of cognition may be based on continuous mappings (MacLennan 1999), rather than on symbol manipulation. Finally, one may inquire as to the form of the laws of cognition that can be expected to arise most naturally from the NS hypothesis. The law of generalization proposed by **SHEPARD** (1987b) is an important first step toward an answer to this question.

***Challenges.*** The two most serious challenges for the NS framework both stem from varieties of holism, albeit rather different ones. First, representing an entire object or event by a point in a neural space precludes the possibility of acting on, or even becoming aware of, its structure (Hummel 2000). Second, the treatment of an object by the cognitive system frequently depends on the context within which the particular problem at hand is situated, and therefore, potentially, on any of the totality of the representations that exist in the system; this observation is used by Fodor (2000) to argue for some very severe limitations on the scope of "computational psychology." It appears to me that both these problems can be addressed within the NS framework. Specifically, adopting a configuration space approach, in which the global representation space approximates the Cartesian product of spaces that code object fragments (Edelman & Intrator 2000), may do away with the unwanted holism in the representation of individual objects. Furthermore, sharing the representation of an object among several neural spaces may support its context-sensitive treatment (as long as the spaces intersect transversely, they can be kept distinct in places away from the intersection). This would allow the system to make the kind of non-compositional, holistic inferences which, as Fodor (2000) rightly notes, abound in human cognition.

NOTES

**1.** "Where instances of a particular collection of intrinsic properties (i.e., properties already diagnosed from sensory information) tend to be grouped such that if some are present, most are, then other useful properties are likely to exist which generalize over such instances. Further, properties often are grouped in this way" (Marr 1970).

**2.** Arguments against this idea based on the observation that perceived similarities can be asymmetrical (Tversky 1977) is effectively countered, for example, by adopting the Bayesian interpretation of Shepard's approach proposed by **TENENBAUM & GRIFFITHS** (this volume). Sticking with the physical space metaphor, one can imagine a foliated, curved neural space, riddled with worm holes (corresponding to arbitrary associations between otherwise unrelated objects or concepts).

## Natural groups of transformations underlying apparent motion and perceived object shape and color

David H. Foster

*Department of Optometry and Neuroscience, University of Manchester Institute of Science and Technology, Manchester, M60 1QD, United Kingdom.*
**d.h.foster@umist.ac.uk** **http://www.op.umist.ac.uk/dhf.html/**

**Abstract:** Shepard's analysis of how shape, motion, and color are perceptually represented can be generalized. Apparent motion and shape may be associated with a group of spatial transformations, accounting for rigid and plastic motion, and perceived object color may be associated with a group of illuminant transformations, accounting for the discriminability of surface-reflectance changes and illuminant changes beyond daylight. The phenomenological and mathematical parallels between these perceptual domains may indicate common organizational rules, rather than specific ecological adaptations.

**[BARLOW; HECHT; KUBOVY & EPSTEIN; SCHWARTZ; SHEPARD; TODOROVIČ]**

***Introduction.*** For the biologically relevant properties of objects such as their position, motion, shape, and color, what sorts of representational spaces offer the possibility of yielding invariant psychological principles? The aim here is to show that the analysis **SHEPARD** uses to address this problem can be generalized. Thus, the phenomenon of rigid apparent motion between sequentially presented objects is cast as a special case of more general kinds of apparent motion and surface-color perception under daylight is cast as an invariant of more general illuminant transformations. Supporting experimental data are cited for each. As a side-effect of this generalization, it may be more difficult to maintain the notion that the rules governing these phenomena are specific adaptations to properties of the world, although they remain illuminating (**SCHWARTZ**, this issue). As with **SHEPARD**'s approach, the present analysis depends critically on choosing appropriate perceptual representations, here based on the natural group structures of the spaces involved.

***Apparent motion and groups of transformations.*** Figure 1a, b shows two possible apparent-motion paths between two sequentially presented bars placed at an angle to each other (adapted from Foster 1978). Of all the possible paths, what determines the one actually perceived? As proposed in Foster (1975b), one way to tackle this problem is to imagine that each path, in some suitable space, has a certain cost or energy associated with it, and, in accord with Maupertuis, the path chosen is the one with least energy. As shown later, energy can be defined in two natural ways: (1) with reference to the space in which the object appears to transform; and (2) with reference to the space of tranformations acting on the object. Neither is a subcase of the other (cf. **KUBOVY & EPSTEIN, TODOROVIČ**, this issue).

How should apparent-motion paths be described? Assume that a stimulus object $A$ and some transformed version of it $T(A)$ are each defined on a region $S$ of some 2- or 3-dimensional smooth manifold constituting visible space. The spatial transformation $T$, which describes the point-to-point relationship between $(A)$ and $T(A)$, should be distinguished from any dynamical process that instantiates this relationship. Depending on the type of apparent motion (rigid or plastic, see Kolers 1972), an object may change its position, its shape, or both. For the sake of generality, therefore, assume that the transformations $T$ are drawn



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

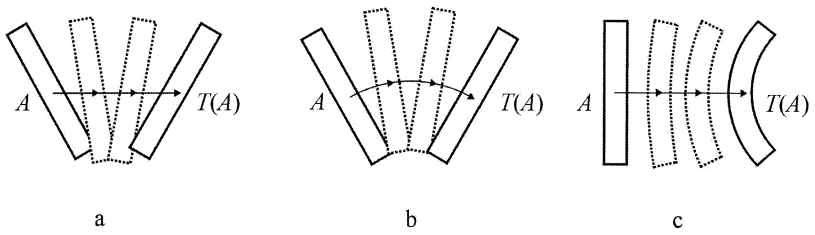



Figure 1 (Foster). Three possible apparent-motion paths in the plane (adapted from Foster 1978).

from a set **T** that is sufficiently large to allow all such possibilities (Foster 1978). For technical reasons, assume also that the space $S$ is compact and connected, and that **T** is a group, with neutral element the identity transformation Id, taking $A$ into itself. Although **T** is a large group, including nonlinear transformations, it is not assumed to coincide with the entire group of diffeomorphisms of $S$.

Apparent motion between $A$ and $T(A)$ can then be represented as the generation by the visual system of a time-parameterized family $c(t)$, $0 \leq t \leq 1$, of transformations defining a path in **T** starting at Id and ending at $T$; that is, $c(0) = \text{Id}$ and $C(1) = T$. (The actual time scale has been set to unity.)

As shown later, the group **T** can be given the structure of a Riemannian manifold, so that at each point $T$ of **T** there is an inner product $\langle , \rangle$ defined on the tangent space at $T$ (the tangent space at a point is simply the collection of all tangent vectors to all possible curves passing through that point). The length $\|v\|$ of a tangent vector $v$ is given by $\langle v, v \rangle^{1/2}$. The length of a path and its (kinetic) energy can then be defined straightforwardly.

***Energy-minimizing paths.*** For each path $c$ in the group **T** of transformations connecting Id to $T$, its arclength $L(c)$ is given by $\int \|c'(t)\| dt$, where $c'(t)$ is the vector tangent to $c$ at $t$ (i.e., the velocity at $c(t)$; see Fig. 2) and the integral is taken over the interval $0 \leq t \leq 1$. If $c$, parameterized by arclength, is not longer than any other path with the same start and endpoints, then $c$ is called a geodesic.

The energy $E(c)$ of $c$ is given by $\int \|c(t)\|^2 dt$, where the integral is again taken over the interval $0 \leq t \leq 1$. It can be shown that the energy $E(c)$ as a function of $c$ takes its minimum precisely on those paths between Id and $T$ that are geodesics. How, then, should the Riemannian metric $\| \ \|$ be defined?

***A natural metric from object space.*** Assume that apparent motion is determined by the properties of the manifold $S$ in which the object appears to transform. As a subset of 2- or 3-dimensional Euclidean space, $S$ inherits the Euclidean metric $| \ |$. For a given object $A$ in $S$, the induced Riemannian metric $\| \ \| = \| \ \|_1$ on **T** is defined thus. Let $c'(t)$ be the vector tangent to a path $c$ in **T** at time $t$ (remember that any tangent vector can be represented in this way). As $c(t)$ is a transformation acting on $S$, it follows that, for each point $p$ in $A$, the vector $(c'(t))(p)$ is tangent to the path $(c(s))(p)$, $0 \leq t \leq 1$, at $s = t$. The energy of $A$ at time $t$ is simply the integral of $|(c'(t))(p)|^2$ over all $p$ in $A$. Define $\|(c'(t)\|_1$ to be integral of $|(c'(t))(p)|$ over all $p$ in $A$.

If **T** is the group of rigid transformations (isometries) of $S$, the geodesics produce the types of motion shown in Figure 1a, where the rotating motion of the bar takes place about its center of mass and the latter moves in a straight line. A matrix formulation is given in Foster (1975b). This is the motion of a free body in space. Yet, as Foster (1975b) and **SHEPARD** point out, it is not the apparent motion that is most likely to be observed.

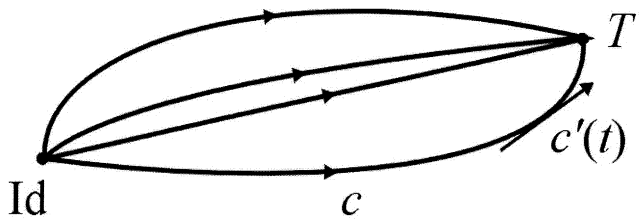



Figure 2 (Foster). Some paths between Id and $T$ in the transformation group **T**.

***A natural metric from transformation space.*** Assume instead that apparent motion is determined by the properties of the group **T** in which the path is described: the emphasis is thus on transformations rather than on transforms. Because **T** is a group, it has a natural Riemannian metric $\| \ \| = \| \ \|_2$, compatible with its group structure, obtained by translating an inner product on the tangent space to **T** at Id. With respect to $\| \ \|_2$, the geodesics $c$ that pass through the Id are (segments of) 1-parameter subgroups of **T**; that is, $c(s + t) = c(s)c(t)$, wherever they are defined.

If **T** is the group of rigid transformations of $S$, the geodesics produce the types of motion shown in Figure 1b, where the rotating motion of the bar and the movement of the center of mass both take place about the same point. A matrix formulation is given in Foster (1975b). When the perceived paths are estimated by a probe or windowing technique, they are found to fall closer to these "group" geodesics than to those associated with object space, namely the free-body motions (Foster 1975b; Hecht & Proffitt 1991; McBeath & Shepard 1989).

***Simplicity of motion.*** **SHEPARD**'s argument for the simplicity of geodesics concentrates on their representation as rotations or screw displacements in the group of rigid transformations of 3-dimensional space. In fact, their simplicity has a more general basis (Carlton & Shepard 1990a; Foster 1975b), which, notwithstanding **TODOROVIČ**, extends to the nonlinear motion shown in Figure 1c between a straight bar and curved bar, and to nonsmooth motion between smooth and nonsmooth objects (Foster 1978; Kolers 1972). To enumerate: (1) group geodesics minimize energy with respect to the natural metric on the group **T**; (2) they coincide with the 1-parameter subgroups of **T**, and are therefore computationally economic in that each may be generated by its tangent vector at the identity Id (Shepard's uniformity principle; see Carlton & Shepard 1990a); and (3) as 1-parameter subgroups each geodesic naturally generates a vector field on $S$ (an assignment of a tangent vector at each point of $S$ varying smoothly from point to point). This assignment does not vary with time; that is, the vector field is stationary. Conversely, a stationary vector field generates a unique 1-parameter subgroup of transformations.

A moving fluid provides a useful example of the significance of stationarity. Its streamlines defined by the velocity vector field usually vary with time, but if the vector field is stationary, then the streamlines are steady and represent the actual paths of the fluid particles.

In general, the geodesics derived from the natural metric of object space (free-body motions) do not generate stationary vector fields.

***Vector fields for generating natural motion.*** The stationarity of vector fields may be relevant to the question of whether kinematic geometry internalized specific properties of the world (**SHEPARD**, this issue). Thus certain vector fields may reflect J. J. Gibson's "ambient optic array" (**SHEPARD**), but they may also relate directly to observers' actions. Some kinds of mental activity, including preparation for movement (Richter et al. 2000) and mental rotation


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

(Deutsch et al. 1988), are associated with neuronal activity in the motor cortex and related areas. Each of the vectors constituting a (stationary) vector field could offer the most efficient template for elementary neural activity to take object $A$ into its transform $T(A)$ (see comments by **BARLOW**, this issue). In this sense, apparent motion might be an internalization not of the ways in which objects move freely in space (cf. **HECHT**, this issue) but of the ways in which observers manipulate or interact with them. Such hypothesis are testable (**BARLOW; KUBOVY & EPSTEIN**, this issue).

***Connections versus metrics.*** The foregoing analysis assumed that the energy of apparent motion is minimized (Foster 1975b; 1978). **SHEPARD**'s approach assumed an affine connection (Carlton & Shepard 1990a). The result, however, is the same.

A connection on any manifold $M$, not necessarily Riemannian, is a rule $\nabla$ that uses one vector field $X$ to transform another vector field $Y$ into a new vector field $\nabla_X(Y)$. Informally, $\nabla_X(Y)$ describes how $Y$ varies as one flows along $X$. In general, even when $X$ and $Y$ commute, $\nabla_X(Y)$ and $\nabla_Y(X)$ need not coincide, but if the connection is symmetric, they do.

A connection provides a sensible notion of parallelism with respect to a path $c$ in $M$. Let $Y(t)$, $0 \leq t \leq 1$, be a parameterized family of vectors such that $Y(t)$ is in the tangent space to $M$ at $c(t)$. Then $Y$ is said to be parallel with respect to $c$ if $\nabla_{c'}(Y)_{c(t)} = 0$ for all $t$. With respect to this connection, a path $c$ is called a geodesic if the family of tangent vectors $c'(t)$ is parallel with respect to $c$.

Now suppose that the manifold $M$ has a Riemannian metric. A connection $\nabla$ on $M$ is compatible with the Riemannian metric if parallel translation preserves inner products; that is, for any path $c$ and any pair $X$, $Y$ of parallel vector fields along $c$, the inner product $\langle X, Y\rangle$ is constant. According to the fundamental theorem of Riemannian geometry, there is one and only one symmetric connection that is compatible with its metric: the Levi-Civita connection.

The geodesics defined as length-minimizing paths in the group $\mathbf{T}$ of transformations are therefore precisely the same as the geodesics defined with respect to the Levi-Civita connection on $\mathbf{T}$. The premiss adopted by **SHEPARD** and Carlton and Shepard (1990) is therefore formally equivalent to that in Foster (1975b).

***Problem of preserving structure.*** A problem with geodesic-based schemes for apparent motion – whether based on metrics or connections – is how to cost the degree to which object structure is preserved. As Kolers (1972) and others have noted, if the rigid transformation $T$ relating two objects is sufficiently large, then the apparent motion may become nonrigid or plastic, even if $T$ has not reached a cut point on the geodesic (e.g., an antipodal point on the sphere).

One way to accommodate this failure is to introduce an additional energy function $E_1$ that represents the cost of preserving metric structure over a path. Such a notion is not implausible. In shape-recognition experiments with stimulus displays too brief to involve useful eye movements or mental rotation, performance is known still to depend strongly on planar rotation angle. Thus, for a rigid transformation $T$ far from Id, the total energy of the geodesic $c$ connecting Id and $T$ would be $\int\|c'(t)\|^2\,dt + E_1(c)$, which could exceed the energy $\int\|b'(t)\|^2\,dt + E_2(b)$ of some other, longer path $b$ connecting Id and $T$, preserving a weaker nonmetric structure with smaller energy function $E_2$.

If this is true, there ought to be a close relationship between apparent motion and visual shape recognition.

***Invariances of motion and a hierarchy of structures for recognition.*** The existence of rigid apparent motion between two objects implies that a visual isometry can be established. In a shape-recognition experiment, therefore, the two objects should be recognizable as each other. This hypothesis has been confirmed for rotated random-dot patterns (Foster 1973). But how should one deal with structures other than metric ones? In practice, one needs a definition of structure that can be interpreted operationally in terms of the transformations (isomorphisms) preserving that structure (Foster 1975a; Van Gool et al. 1994). For (1) metric, (2) affine, (3) projective, and (4) topological structures, their groups of isomorphisms form a nested sequence, $\mathbf{T}_1 \subset \mathbf{T}_2 \subset \mathbf{T}_3 \subset \mathbf{T}_4$. Accordingly, for one of these more general structures $i$, suppose that transformation $T$ is drawn from $\mathbf{T}_i$ and that sequentially presenting object $A$ and transform $T(A)$ produces apparent motion that lies entirely within $\mathbf{T}_i$. Then, in a shape-recognition experiment, $A$ and $T(A)$ should be recognizable as each other with respect to the structure $i$. Such an exercise offers the possibility of identifying an underlying structure for visual space (Foster 1975a; Indow 1999).

The remainder of this commentary is concerned with perceived surface color, the analysis of which has parallels with the analysis of apparent motion.

***Surface color.*** The illumination on surfaces varies naturally, and the spectrum of the light reaching the eye depends both on the reflectance function of the surface and on the illuminant spectrum. **SHEPARD** suggests that the intrinsically 3-dimensional nature of daylight is intimately linked to how observers compensate for illuminant variations.

Yet the degree to which observers are color constant is limited, with levels in the unadapted eye rarely exceeding 0.6–0.7, where on a 0–1 scale, 1 would be perfect constancy (for review, see Foster et al. 2001). In contrast, observers can rapidly, effortlessly, and reliably discriminate illuminant changes on a scene from simultaneous changes in the reflecting properties of its surfaces (Craven & Foster 1992). The sequential presentation of the stimuli generates a strong temporal cue: illuminant changes give a "wash" over the scene and reflectance changes a "pop-out" effect (Foster et al. 2001). The former is analogous to apparent motion between an object and its smooth transform, and the latter to split apparent motion between an object and its discontinuous transform.

If perceived surface is not always preserved under illuminant changes, then what is invariant in discriminations of illuminant and material changes?

***Invariance of spatial color relations.*** One possibility is that observers assess whether the perceived relations between the colors of surfaces are preserved, that is, whether relational color constancy holds. Relational color constancy is similar to color constancy but refers to the invariant perception of the relations between the colors of surfaces under illuminant changes. It has a physical substrate in the almost-invariant spatial ratios of cone excitations generated in response to light, including illuminants with random spectra, reflected from different illuminated surfaces (Foster & Nascimento 1994). There is strong evidence that observers use this ratio cue, even when it may not be reliable (Nascimento & Foster 1997).

In the language of geometric-invariance theory, relational color constancy is a relative invariant with respect to illuminant changes, and in that sense, is a weaker notion than color constancy (Maloney 1999). But relational color constancy can be used to produce color-constant percepts. Again, the argument depends on group properties.

***Groups of illuminant transformations and color constancy.*** The set $\mathbf{T}$ of all illuminant transformations $T$ is a one-to-one copy of the multiplicative group of (everywhere-positive) functions defined on the visible spectrum, and it accordingly inherits the group structure of the latter. The group $\mathbf{T}$ induces (Foster & Nascimento 1994) a canonical equivalence relation of the space $\mathbf{C}$ of all color signals (each signal consisting of the reflected spectrum at each point in the image). That is $C_1$ and $C_2$ in $\mathbf{C}$ are related if and only if $T(C_1) = C_2$ for some $T$ in $\mathbf{T}$.

The assumption of color constancy is that it is possible to find some $f$ that associates with each $C$ in $\mathbf{C}$ a percept $f(C)$ that is invariant under illuminant transformations. Because $\mathbf{T}$ is a group, there is a one-to-one correspondence between color-constant percepts $f(C)$ and equivalence classes $[C]$ of illuminant-related color signals. This formal equivalence between color constancy and relational color constancy can be exploited in practical measurements (e.g., Foster et al. 2001).

As **SHEPARD** points out, although we may not perceive everything that could be perceived about each surface, we at least perceive each surface as the same under all naturally occurring con-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ditions of illumination, and as argued here, sometimes even under unnatural illuminants.

***Summary and conclusion.*** The representations of apparent motion and perceived shape and object color are intimately associated with groups of spatial transformations. In **SHEPARD**'s analysis, the geodesics for apparent motion are attributed to an affine connection, but the same geodesics can be derived as the natural energy-minimizing paths of a transformation group, which allows an additional energy function to be introduced to accommodate rigid-motion breakdown, and more generalized kinds of shape recognition. In SHEPARD's analysis of perceived object color, daylight illuminants have a special role, but the same perceptual invariants may be obtained with a group of illuminant transformations taking illuminants beyond the daylight locus.

What of the evidence? For rigid transformations in 2- and 3-dimensional space there is a clear bias toward motions following the natural transformation-group metric. There is also evidence that rigid apparent motion does not occur at angles of rotation where shape recognition does not occur, consistent with the proposed link between the two phenomena. Finally, there is evidence that observers can exploit violations of invariance of spatial color relations under illuminant transformations in a predictable way.

The phenomenological and mathematical parallels between these various perceptual domains may not be consequences of **SHEPARD**'s notion of adaptation to specific properties of the world. They do, however, suggest an application of common organizational rules.

ACKNOWLEDGMENTS

This work was supported by the Engineering and Physical Sciences Research Council. I thank E. Pauwels for helpful discussions and E. Oxtoby for critically reading the manuscript.

## Interpreting screw displacement apparent motion as a self-organizing process

T. D. Frank, A. Daffertshofer, and P. J. Beek

*Faculty of Human Movement Sciences, Vrije Universiteit, Amsterdam, 1081 BT, The Netherlands.* **t.frank@fbw.vu.nl marlow@fbw.vu.nl p_j_beek@fbw.vu.nl http://www.fbw.vu.nl/**

**Abstract:** Based on concepts of self-organization, we interpret apparent motion as the result of a so-called non-equilibrium phase transition of the perceptual system with the stimulus-onset asynchrony (SOA) acting as a control parameter. Accordingly, we predict a significantly increasing variance of the quality index of apparent motion close to critical SOAs. **[SHEPARD]**

In his target article, **SHEPARD** demonstrates that apparent motion is characterized by a few basic features. We interpret these features in terms of the theory of self-organization in complex systems (e.g., Haken 1977; Nicolis 1995), which has been successfully applied in a wide variety of fields, ranging from laser physics (Haken 1985) to studies of human movement (Beek et al. 1995; Haken 1996; Kelso 1995) and perception (e.g., Ditzinger & Haken 1989; 1990; Fukushima 1980; Haken 1991; Haken & Stadler 1990; Riesenhuber & Poggio 2000) including studies of orthogonal and circular apparent motion (Hock et al. 1993; Kruse et al. 1996). According to this theory, the interactions among the constituent parts of a self-organizing system may give rise to a complete set of time-independent spatial modes. For apparent motion we can assume that these modes describe all possible types of transformation (e.g., rotations, translations) connecting two alternately displayed object-presentations. Of particular empirical importance is the screw displacement mode, which can be described by an infinitely long, 360 degrees wide strip. Recognizing the periodicity of the rotational transformation, we may roll up this strip. The resulting cylinder-shaped screw displacement mode can then be viewed as analogous to the manifolds discussed in the target article.

Having a complete set of transformational modes at our disposal, we can express a response of the perceptual system to the alternately displayed object-presentations in terms of a superposition of these modes, where each mode is weighted by a time-dependent coefficient. The relevance of each mode can be evaluated when qualitative changes of the perceptual system are observed, for example, close to critical SOAs. Expressed in the parlance of the theory of self-organization, the observed qualitative changes in perceptual experience (rigid vs. nonrigid motion) are identified as non-equilibrium phase transitions and the SOA as the control parameter inducing these transitions. In the immediate vicinity of such transitions, the transformational modes can be classified into stable and unstable modes. Importantly, the evolution of the stable modes can be expressed in terms of the time-dependent coefficients of the unstable modes (Haken 1977). Hence, the latter coefficients are identified as so-called order parameters. As a rule, there are only a few unstable modes and, thus, only few order parameters. Assuming a single order parameter, the response of the entire system is solely dominated by the evolution of just this order parameter (in conjunction with the operations specified by the corresponding unstable mode). In the case of multiple competing order parameters, one order parameter may survive while the others may die out (e.g., Haken 1977; 1991). The "winner" and its concomitant unstable mode will then again govern the system's behavior. In view of the experimental evidence presented in the target article, one can conclude that apparent motion usually exhibits only a single unstable mode, that is, the screw displacement mode. For almost symmetrical objects, however, competition between order parameters related to different unstable modes has also been observed (Farrell & Shepard 1981).

Focusing on the evolution of the (winner) order parameter, its instantaneous value is described by a point on the cylindrical surface of the screw displacement mode. A screw displacement is characterized by a constant ratio of rotational and translational displacements. Put differently, the order parameter evolves along a geodesic. According to the geodesic hypothesis (Carlton & Shepard 1990a), one may stress that geodesics are defined as the shortest paths between two points on a given surface. In search of a neural mechanism leading to geodesically-curved order parameter trajectories, one may assume that the act of connecting two different object-presentations by means of a spurious path requires computational efforts which increase with the length of the path. This hypothesis can be supported by experimental findings regarding the imagination of object rotations, where performance time increases with the angular disparity of the initial and target orientations (Shepard & Metzler 1971). Similarly, the tendency of the perceptual system to minimize its effort is also observed in human and animal locomotion, where gait patterns are selected so as to minimize energy consumption (e.g., Hoyt & Taylor 1981; Minetti & Alexander 1997).

In fact, with the a priori existence of a complete set of transformational modes, there is no need for any internalization of the screw displacement mode itself. Just as visual hallucination patterns (Ermentrout & Cowan 1979) and movement-related patterns of brain activity (Frank et al. 1999), the transformational modes emerging in the apparent motion can be assumed to arise from the interaction of inhibitory and excitatory neurons of the visual system under particular boundary conditions. In our opinion, the issue of internalization arises when we attempt to understand why a particular mode becomes unstable and starts to dominate the system's behavior. Following **SHEPARD**'s considerations, we may say that experiences such as the shape conservation of solid objects or the performance of the screw movements induce inhomogeneities in the neural perceptual system. These may affect, in particular, transformational modes related to everyday experiences like the screw displacement mode.

It follows from the theory of non-equilibrium phase transitions that the variance of the quality index used by Shepard and Judd (1976) should increase significantly when the SOA approaches its critical value. Such (critical) random fluctuations have been observed in coordinated rhythmic movements (Kelso et al. 1986) and



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

in movement-related brain signals (Frank et al. 1999; Wallenstein et al. 1995). In addition, the phenomena of hysteresis (Farrell & Shepard 1981) and multistability (here in terms of multiple apparent motion paths of L-shaped objects) is at the core of our understanding of self-organizing systems (Daffertshofer et al. 1999; Frank et al. 2000; Haken 1996; Hock et al. 1993; Kruse et al. 1996; Peper et al. 1995). Therefore, further experimental studies on apparent motion elaborating on these features seem rewarding.

ACKNOWLEDGMENT
David Jacobs is sincerely thanked for stimulating discussions.

## Exhuming similarity

Dedre Gentner
*Psychology Department, Northwestern University, Evanston, IL 60208.*
**gentner@northwestern.edu**
**http://www.psych.northwestern.edu/psych/people/faculty/gentner/**

**Abstract:** Tenenbaum and Griffiths' paper attempts to subsume theories of similarity – including spatial models, featural models, and structure-mapping models – into a framework based on Bayesian generalization. But in so doing it misses significant phenomena of comparison. It would be more fruitful to examine how comparison processes suggest hypotheses than to try to derive similarity from Bayesian reasoning.
**[SHEPARD; TENENBAUM & GRIFFITHS]**

**TENENBAUM AND GRIFFITHS'** (**T&G**'s) paper is large in its vision. It aims to synthesize similarity, concept learning, generalization, and reasoning. Under the rubric of Bayesian generalization, it offers a unification of **SHEPARD**'s spatial model of similarity with Tversky's set-theoretic model and, for good measure, with structure-mapping accounts of similarity. This is a bold and ambitious idea, but there are some problems.

***Getting directionality correct.*** First consider the unification of Tversky's (1977) contrast model with SHEPARD's spatial model and with generalization. **T&G** give Tversky's statement of the similarity of a target y to a base x:

$$S(y,x) = \Theta f(Y \cap X) + \alpha f(Y - X) - \beta f(X - Y)$$

To achieve a unification of Tversky's contrast model with their Bayesian model of generalization, T&G first assume the special case of the contrast model in which the measure f is additive. Then they adopt the ratio version of the contrast model, and finally they set $\alpha = 0$ and $\beta = 1$. QED – at this point the similarity equation does indeed resemble the Bayesian equation. But these weightings of $\alpha$ and $\beta$ are the reverse of Tversky's. In Tversky's model, $\alpha > \beta$; that is, the distinctive features of the target term (y) count more against the similarity of the pair than the distinctive features of the base (or standard, or referent) term (x). This turns out to matter. This is how Tversky explains the finding that people think that Nepal is more similar to China than the reverse. For most of us, China is the richer concept. When it is in the base position, its distinctive features get a lower weight ($\beta$) so the similarity of the pair is greater. The T&G formulation predicts directionality preferences that are the reverse of what's normally found.

***Common relations are not scorned.*** **T&G** also take up structural models of similarity (Gentner & Markman 1997; Medin et al. 1993). They note that people in our studies often find common relations more important than common object attributes (primitive features) in similarity judgments. For example, people typically consider AA to be more similar to BB than to AC. T&G propose that the relational preference can be derived from Bayesian principles – specifically, from the size principle, that people prefer hypotheses that delineate small consequential regions – that is, that they prefer specific hypotheses to general hypotheses. (We might note in passing that the assumption that specific hypotheses are superior to general ones is not unique to Bayesian theories.) **T&G** conjecture that the greater saliency of relations over primitive object features stems from their greater specificity. Because relations are more specific than objects, a generalization based on relations is more informative. T&G use this rarity principle to offer an explanation for why *same* relations are more salient than *different* relations.

But the rarity explanation for why people pay attention to relations runs into some immediate difficulties. First, in analogy and similarity, higher-order relations such as causal relations are extremely highly weighted, despite being ubiquitous in human reasoning. Second, in natural language descriptions people tend to use relational terms broadly and object terms specifically: that is, they use a relatively small number of high-frequency relational terms, each very broadly, and a large number of low-frequency object terms, each quite specifically. This suggests that it is *objects* that are specific, not relations. To test the claim that the relational preference in comparison does not depend on rarity or specificity, I asked people for similarity judgments in triads that had low-frequency, specific object terms and high-frequency, rather general relational terms. Given the triad

Blacksmith repairing horseshoe

Blacksmith having lunch (A) Electrician repairing heater (B)

15 out of 19 people chose the relational response (B) as most similar to the standard, despite the fact that *repairing* is broadly applicable and frequently encountered, and *blacksmith* is highly specific and rarely encountered.

Another problem with the general claim that relations are salient relative to object features because of their specificity is that relations *aren't* always more salient. Relations have a salience advantage when comparing two present terms (Goldstone et al. 1991), but object features are more salient than relations in similarity-based memory *retrieval* (Gentner et al. 1993; Ross 1989). This disassociation between the relative salience of objects and relations during mapping versus during retrieval cannot be accommodated with an all-purpose salience assignment.

There are other problems with subsuming similarity under generalization, such as that comparison processes systematically highlight not only commonalities but also certain differences – alignable differences, those connected to the common system. **T&G**'s aim of extending classic models of similarity to structured stimuli is laudable. But the attempt to subsume similarity under Bayesian generalization fails to capture some basic phenomena. Moreover, a processing account of comparison could provide one of the missing elements in Bayesian theories, an account of how people arrive at their hypothesis spaces.

ACKNOWLEDGMENT
I thank Doug Medin, Lera Boroditsky, and Art Markman for discussions of this paper.

## The place of Shepard in the world of perception

Walter Gerbino
*Department of Psychology, University of Trieste, 34134 Trieste, Italy.*
**gerbino@univ.trieste.it** **http://www.psico.univ.trieste.it/**

**Abstract:** To balance Kubovy & Epstein, I evaluate the relationship between Shepard and Gestalt theorists along three dimensions. First, both discover internal universals by reducing external support. Second, they share strengths and weaknesses of the minimum principle. Third, although their attitudes toward an evolutionary account of perception is superficially different, they are fundamentally similar with respect to the internalization process.
**[KUBOVY & EPSTEIN; SHEPARD]**

**KUBOVY & EPSTEIN** (K&E) locate **SHEPARD** "in the neighborhood staked by Helmholtz and Rock" although SHEPARD has formerly



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

"aligned himself with Helmholtz's stance" and "resonated to Gibson's resonance theory" (sect. 1.2, "Locating Shepard"). **K&E**'s theoretical landscape includes Helmholtz, Transactionalism, Rock, Marr, and Gibson, but not Gestalt theory. SHEPARD contributed to the Gestalt revival (see his chapter in Kubovy & Pomerantz 1981) and likes the minimum principle (Hatfield & Epstein 1985). Therefore K&E should have good reasons not to tell us where they see him in the perception valley, relative to Gestalt monuments. I suggest that SHEPARD is very close to Gestalt theory (sects. 1, 2) and only deceptively far along the evolutionary dimension (sect. 3).

***1. External support vs. internal universals.*** Apparent motion paths "experienced in the absence of external support are just the ones that reveal, in their most pristine form, the internalized kinematics of the mind and, hence, provide for the possibility of an invariant psychological law" (**SHEPARD**'s sect. 1.7, "Conditions revealing the default paths of mental kinematics"). **K&E** remark that this is a standard procedure (beginning of sect. 2.2, "Questioning internalization"). However, exploring percepts as end-products of the equilibrium between external and internal forces has been *the* Gestalt strategy (Koffka 1935, chapter IV). Metzger (1941/1954, chapter VI, sect. 7) emphasized that the weakening of external forces reveals different regularization phenomena, corresponding to deviations from stimulus properties, dimensionality, or articulation. SHEPARD's work on apparent motion disclosed internal principles of spatiotemporal interpolation, parallel to those involved in the spatial interpolation of amodally completed contours, surfaces, and volumes (Gerbino 1997; Kanizsa & Gerbino 1982; Kellman & Shipley 1991; Singh & Hoffman 1999; Tse 1999a; 1999b).

***2. Simplicity.*** **SHEPARD** and Gestalt theory share an important distinction between formal (e.g., minimal coding) and processing simplicity.

> [W]hen a simple screw displacement or rigid rotation is possible, that motion will tend to be represented because . . . it is the geometrically simplest and hence, perhaps, the most quickly and easily computed. Certainly, . . . such a motion requires the minimum number of parameters for its complete specification. (SHEPARD's sect. 1.4, "Kinematic simplicity is determined by geometry")

However, they also share a theoretical weakness about the choice of *appropriate* parameters for defining simplicity, singularity, regularity, order, prägnanz, minimum principle, and related notions.

> This single rigid rotation is geometrically simpler than the motion prescribed by Newtonian mechanics, which generally includes two components: a continuous motion of the center of mass (which is rectilinear in the absence of external forces), and an independent rotation about that moving center. (SHEPARD's sect. 1.5, "Geometry is more deeply internalized than physics")

It is not clear why, on formal grounds, two differently oriented views of a 2D shape are *better* interpolated by a rigid rotation around an external point (the perceptually preferred solution) than by a translation of the center combined with a rotation of the shape around it. If rigid planar motions are decomposable into rotations (and translations are taken as rotations around a point at infinity), there are two component motions: (*a*) rotation of center C around point P, involving two parameters, CP *length* and *angular extent;* (*b*) rotation of the shape around its center C, involving only *angular extent.* The kinematics of any path requires the specification of three values. When a given value is zero, one *cannot* disregard the corresponding parameter and treat it as irrelevant. One might argue that the system tends to minimize parameter values (not only the number of parameters). However, here is the shared theoretical weakness: the cost function for value minimization is not specified and researchers tend to claim that a solution is formally simpler *after* discovering that it is preferred by perception (or imagery or naive reasoning). SHEPARD considers 2D-motions involving three non-zero values as more complex than the two motions involving one zero value (for the angular extent of rotation *b* or rotation *a,* respectively), which means that he has in mind a cost function with uneven weights for the three parameters. Otherwise the simplest motion would correspond to an optimal combination of three non-zero values. But claiming that *the best* solution is the rotation around an external point combined with zero rotation around the center of the shape, requires the assumption that motion of the shape relative to its intrinsic framework (orientation axes centered in C) is *more* important than motion relative to the extrinsic framework (orientation axes centered in P). It might be so (Koffka 1935, chapter VI; Metzger 1941/1954, chapter IV), but the problem of hierarchical motion frameworks must be clarified.

***3. Internalization as a phylogenetic process.*** The visual system *follows* or *instantiates* internal principles that correspond to the external world (**K&E**'s sect. 1.1, "Precursors"). Such a correspondence or complementarity (Shepard 1981b) is a well-established *fact* that can be called *internalization.* However, the same term (see K&E's Abstract) can stand for a *process* based on natural selection over the evolutionary history of the species. K&E argue that the internalization process is a mere metaphor we can live without, a metaphor that should not be confused with the material intake of external things. I fully agree. But in the immaterial world of information processing, the eating-digesting metaphor is common (Kolers 1972; Ramachandran 1990b) and I am confident that over the years scientists have been wise enough not to explore its extreme implications. Furthermore, the point is *not* whether we can live without such a metaphor (of course we can); rather, whether life is better with such a metaphor. Evolutionary-oriented arguments provide an attractive point of view and **SHEPARD**'s genes that shape perceptual and cognitive capabilities have inspired great experiments. However, I admit that on this matter SHEPARD can be located very far from Gestalt theory. Koffka (1935, chapter XIII) viewed biologizing as a way of explaining away the problem of internal universals and falling into "the trap of a teleological explanation."

> To say: a certain process occurs because it is biologically useful, would be the kind of explanation we have to guard against. For the biological advantage of a process is an effect which has to be explained by the process, but the former cannot be explained by the latter. A process must find its explanation in the dynamics of the system within which it occurs; the concept of biological advantage, on the other hand, does not belong to dynamics at all. And therefore teleological explanations in terms of biological advantage have no place in gestalt theory. (Koffka 1935, chapter XIII)

Köhler (1929, chapter V) took for granted "the enormous biological value of sensory organization," which "tends to have results which agree with the entities of the physical world," and provided a clear formulation of vision as inverse optics: "In countless instances sensory organization means a reconstruction of such aspects of physical situations as are lost in the wave messages which impinge upon the retina." A balanced Gestalt view of the relationship between evolution and causation is the following:

> This need [*taking an evolutionary point of view*] is right, because the study of . . . a new organic event or behavior . . . begins with the question of the goals they fulfill . . . . But also in biology, when the correct answer to such a question is found, the other question . . . about the internal and external causes or conditions of the event, remains unanswered. The second question does not become irrelevant. On the contrary, it remains equally if not more mysterious. When the question about the goal is answered, the question about conditions for such successful achievements becomes even more urgent. The main scientific task only starts here. (Metzger 1941/1954, chapter VII, sect. 6.1)

Regarding the evolutionary dimension (sect. 3), **SHEPARD** could be perceived as being far from Gestalt theorists, but like them he takes for granted the biological value of perceptual organization, despite our limited knowledge of *how* organizing principles become internalized. Although the internalization process is not well understood, I would not claim that "it has no obvious empirical content and cannot be tested experimentally" (**K&E**'s sect. 2.2, "Questioning internalization"). Internalization can be simulated and the survival value of acquired behavioral patterns can be tested. However questionable current simulations of perceptual


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

adaptation are, there are no *a priori* reasons for rejecting them as invalid tests of internalization, provided they are used to compare systems that evolved in different worlds. Such a requirement might put outside the reach of empirical testing SHEPARD's kinematics (as argued by K&E), but not all internal principles.

## The evolution of color vision

Ian Gold

*Department of Philosophy, Monash University, Clayton VIC 3168, Australia.*
**ian.gold@arts.monash.edu.au**

**Abstract:** It is argued that color constancy is only one of the benefits of color vision and probably not the most important one. Attention to a different benefit, chromatic contrast, suggests that the features of the environment that played a role in the evolution of color vision are properties of particular ecological niches rather than properties of naturally-occurring illumination.
**[SHEPARD]**

Color vision is an adaptation that is widespread across diverse species and is, therefore, a promising area in which to look for confirmation of **SHEPARD**'s hypothesis of the internalization of physical universals. SHEPARD's argument concerning color runs as follows. Color vision is adaptive because color constancy is useful to animals. Color constancy can be achieved by internalizing principles concerning the three dimensions of naturally-occurring illumination and compensating for variations along those dimensions. The internalized principles concerning the illumination make their appearance in cognition as the three dimensions of opponent color appearance space. In this commentary, I challenge SHEPARD's assumption that color constancy is the central adaptation of color vision, and I argue that focusing on a different benefit of color vision undermines SHEPARD's views about the universality of cognitive principles.

It is certainly plausible that color constancy is adaptive. For example, the ability to identify a predator by its color whatever the ambient lighting conditions is likely to be useful to an organism. But this capacity is not the only – and probably not the most important – benefit of color vision. Walls (1942, p. 463) long ago pointed out that chromatic contrast increases the visibility of objects dramatically. A tiger that reflects the same amount of light as the surrounding foliage will be largely invisible to an animal that has only brightness contrast. The contrast between the tiger and the foliage is enormously enhanced, however, if the light reflected by each comes from a different part of the visible spectrum, and the perceiver can discriminate wavelengths. One of the benefits of color vision, therefore, is that it makes it possible to see more than one could without it. Indeed, it is plausible that this aspect of color vision is more important than seeing objects that are *already* visible as having the same color across differing illuminations. For this reason, it is likely that the evolution of color vision was driven by the advantages of contrast at least as much as by constancy, and probably more so. Further, because chromatic contrast requires only wavelength sensitivity, but constancy requires something more, it is likely that contrast evolved first.

Attention to chromatic contrast, however, tends to highlight the significance of the physical properties of particular ecological niches as against the global properties of the illumination because the contrasts of importance to an animal will be those that carry informational significance within that niche. Trichromacy in Old World primates, for example, may have evolved to facilitate the detection of fruit against green leaves (Osorio & Vorobyev 1996), and the tuning of human photopigments can be interpreted in the light of this suggestion (Osorio & Bossomaier 1992). Bees are particularly interested in flower color and may have evolved sensitivity to ultraviolet (UV) light because some flower patterns are only visible under UV (Menzel & Backhaus 1991; see also Dennett 1991). Other birds, including domestic chicks also detect UV light (Vorobyev et al. 1998), and the UV-sensitive cones of chicks form part of an opponent mechanism (Osorio et al. 1999). Butterflies use color vision both for the identification of food and conspecifics, and the distribution of visual pigments differs in the male and female of the species according to the differing behaviors of the two sexes (Bernard & Remington 1991).

Supporting evidence for the local nature of color evolution comes from the fact that color vision systems vary across different species even within primates (Jacobs 1996; Jacobs et al. 1996). Further, the opponent color spaces of some species may not exhibit the same structure as that of humans. A description of the color judgements of honey bees, for example, does not require positing a dimension of brightness even though bees can discriminate brightness well in certain behavioral circumstances (Menzel & Backhaus 1991). Although it is difficult to infer the function of color systems and color discrimination behavior from physiology, some facts, such as the differential tuning of photopigments, constitute prima facie evidence that there are functional color differences in different species.

Therefore, if contrast is a significant aspect of color vision, it is likely that the features that drove its evolution are local features of the ecological niches of particular species and not the global properties of the illumination. Even if contrast is only one among a number of benefits that come with color vision, an evolutionary account of chromatic contrast will not lead to the positing of universals as **SHEPARD**'s account of the evolution of color contancy does. Thus, even if SHEPARD is correct in positing the existence of internalized principles that facilitate the perception of color, at least some of these principles are likely to be specific to particular species and niches rather than uniform across all animals that perceive color.

## What are we talking about here?

John Heil

*Department of Philosophy, Davidson College, Davidson, NC 28036, USA and Department of Philosophy, Faculty of Arts, PO Box 11a, Monash University, Clayton, Vic. 3800, Australia.* **joheil@davidson.edu**
**john.heil@arts.monash.edu.au**
**http://www.davidson.edu/academic/philosophy/joheil.html/**

**Abstract:** Shepard provides an account of mechanisms underlying perceptual judgment or representation. Ought we to interpret the account as revealing principles on which those mechanisms operate or merely an account of principles to which their operation apparently conforms? The difference, invisible so long as we remain at a high level of abstraction, becomes important when we begin to consider implementation.
**[SHEPARD]**

What should we ask of a science of psychology? Roger **SHEPARD** provides one kind of answer. Psychological explanation aims for the mathematical simplicity and elegance, if not the precision, of the best physical sciences. Terrestrial creatures (and presumably extra-terrestrial counterparts) have evolved to survive and flourish in dynamic, ecologically diverse environments. Creatures so evolved could well have availed themselves of underlying, mathematically tractable regularities exhibited by their surroundings, regularities painstakingly exposed by the physical sciences. Suppose, for instance, we look carefully at information available to creatures' perceptual systems and work out principles that could take us from those inputs to perceptual judgments. We should not be satisfied with analyses that merely happen to fit the data. On the contrary, we should look for commonality across cognitive domains. With luck we might uncover simplicity and unity underlying superficially diverse phenomena. Principles invoked to account for perceived color constancies might, for instance, share important mathematical properties with principles required to explain perceived shape constancies.

These are laudable aspirations; and **SHEPARD**'s longstanding pur-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

suit of them has yielded impressive results. A question remains, however, one that falls outside the province of Shepard's discussion. The question concerns what exactly Shepard's equations purport to describe. Shepard speaks of "representations" and "appearances." This implies that what Shepard is after is a set of principles governing creatures' manipulation of representational states of mind.

Physicists employ equations to represent and explain the actions and powers of material bodies: bodies "obey" or "follow" laws these equations express. In just the same sense, intelligent creatures could be said to obey or follow laws of physics. Compare this to a case in which you obey a rule for stopping at stop signs by halting your car at a stop sign. Invoking a distinction made famous by Kant, we can say that you are guided by or *act on* the stop sign rule. In contrast, although your actions *accord with* laws of nature, you do not act on those laws. This is just to say that actions can accord with a law or principle without thereby being based on or guided by that law or principle. In acting on a principle, an agent's grasp or representation of the principle (in concert with other states of mind) controls the action.

What of **SHEPARD**'s principles? Suppose Shepard has it right: creatures' assessments (explicit or implicit) of certain features of their environment conform to the principles he advances. Do these principles guide creatures' assessments of colors, or shapes, or motions? That is one possibility. Another, less ambitious, possibility is that creatures' actions merely satisfy the principles. If that were so, then the creatures need only possess a nature the physical composition of which supports mechanisms whose operation is describable via the principles.

Is this one of those philosophical distinctions without a difference? Certainly anything any creature does, if governed by any law, is governed by (and so accords with) basic physical law. Just as this need not be taken to imply that every science is reducible to (in the sense of being replaceable by) physics and chemistry, so it need not mean that explanations that appeal to principles on which agents are taken to act, are replaceable by explanations framed in terms of laws to which agents' actions merely conform. In invoking representations in explanations of creatures' actions we appeal to this very distinction.

Representing our surroundings differs from simply mirroring those surroundings. Representation is selective and partial; we represent the world in a particular way from a particular point of view. Evolution ensures that perceptual representations are constrained by the world. Our finite nature imposes additional constraints. This can be made to sound trite: the way the world looks, feels, sounds, and tastes to us depends on how the world is and how we are. But the formulation of principles that capture the workings of this mechanism is anything but trite.

Psychological explanation is susceptible to a peculiar sort of mis-direction. Features of the explanatory apparatus are easily mistaken for features of what is being explained. This occurs in everyday life when we anthropomorphize pets, ascribing to them states of mind they are unlikely to be in a position to harbor. Psychologists risk a similar confusion in formulating principles taken to govern mental representations. It is easy to mistake features of the formulation for features of the system. An example of a mistake of this kind might be the imputation of a mechanism for solving differential equations in the brain of an outfielder pursuing a fly ball. We describe the ball's trajectory using differential equations, and the outfielder's brain must incorporate mechanisms that arrive at comparable solutions. But it need not follow that the brain engages in computations of the kind we would use to describe the flight of the ball. Instead, the brain might avail itself of simpler heuristic mechanisms. One way to describe these mechanisms is to describe their inputs and specify a principle that takes these into appropriate outputs. But we cannot move directly from such a description to the conclusion that the brain operates on, and not merely in accord with, these principles.

Perhaps the nervous system is a "smart machine" or, better, a vast system of smart machines (Heil 1983; Runeson 1977). Smart machines are devices that execute computationally sophisticated tasks in mechanically simple ways. A centrifugal governor on a steam engine is a smart machine, as is a polar planimeter (a simple device used to determine the area of irregular spaces, the area of an island on a map, for instance). Such devices act in accord with certain mathematical rules, but not on the basis of those rules. Knowing the rules would not tell you how the devices were constructed, how they actually operate. It is hard to avoid the impression that **SHEPARD**'s principles are like this. In representing the world, we (or our visual systems) act in accord with these principles, but not on them. This is where talk of mental representations stands to be misleading. Mechanisms underlying the production and manipulation of our worldly representations could well operate in accord with certain principles without those principles mirroring the underlying mechanisms. Mechanisms operating in accord with the very same principles could well differ internally in important ways.

None of this affects the validity or significance of **SHEPARD**'s results – results which, in any case, a philosopher is in no position to challenge. It does, however, affect the ways we might seek to understand and test those results in looking at the underlying hardware.

ACKNOWLEDGMENTS

The author is indebted to Davidson College for funding a research leave during 2000–2001, and to the Department of Philosophy, Monash University, for its hospitality, intellectual stimulation, and support.

## What is the probability of the Bayesian model, given the data?

Evan Heit

*Department of Psychology, University of Warwick, Coventry CV4 7AL, United Kingdom.* **E.Heit@warwick.ac.uk**
**http://www.warwick.ac.uk/staff/E.Heit/**

**Abstract:** The great advantage of Tenenbaum and Griffiths's model is that it incorporates both specific and general prior knowledge into category learning. Two phenomena are presented as supporting the detailed assumptions of this model. However, one phenomenon, effects of diversity, does not seem to require these assumptions, and the other phenomenon, effects of sample size, is not representative of most reported results. **[TENENBAUM & GRIFFITHS]**

The Bayesian model proposed by **TENENBAUM & GRIFFITHS** has a number of strengths, such as extending Shepard's (1987b) account of generalization to multiple stimuli. This model is by no means the only model of categorization that extends **SHEPARD**'s work (see, e.g., Nosofsky 1988b), nor is it the only Bayesian model of categorization to be applied to psychological data (see, e.g., Anderson 1991). Perhaps what is most important and novel about this modeling effort is the explicit emphasis on how people's prior beliefs are put together with observed category members to make classification judgments.

Since Murphy and Medin (1985), there have been many theoretical arguments and empirical demonstrations showing that categorization must be constrained by prior knowledge and cannot simply depend on generalization from observations (see Heit 1997b, for a review). However, model-based research in categorization has lagged behind on this important issue, with most categorization models not addressing influences of prior knowledge. In contrast, the Bayesian model of **TENENBAUM & GRIFFITHS** gives an elegant account of how two kinds of prior knowledge are incorporated into categorization.

First, category learning is set against the backdrop of a hypothesis space, which represents expectations about the possible content of the category. Category learning can be viewed as elimination of hypotheses that do not fit the data while strengthening the remaining hypotheses (cf., Horwich 1982). The Bayesian method for deriving posterior probabilities of hypotheses embodies the idea that not only does prior knowledge serve as a guide to what



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

the observed category members will be like, but also the observations themselves are crucial for selecting from among numerous prior hypotheses (Heit & Bott 2000).

Second, the modeling framework can apply general knowledge about how observations are sampled. This is knowledge not about the possible content of the category to be learned, but rather about the manner of learning itself. The crucial idea introduced by **TENENBAUM & GRIFFITHS** is "strong sampling," an assumption that observations are drawn randomly from some fixed population. Strong sampling has important consequences, such as favoring specific hypotheses corresponding to smaller populations of positive examples – this is called the "size principle." Within this modeling framework it could be possible to build in further distinctions about sampling, such as whether sampling is with or without replacement (Barsalou et al. 1998) or whether the observations have been presented in some purposeful order according to goals of a teacher (Avrahami et al. 1997).

In support of the Bayesian model including the size principle, **TENENBAUM & GRIFFITHS** focus on two phenomena, that more variable or diverse observations lead to broader generalizations, and that as the number of observations within a given range increases, generalization outside the range is reduced. These two phenomena are now considered in turn.

First, although the effect of diversity does appear to be robust, there have been salient exceptions reported in inductive reasoning tasks (reviewed by Heit 2000). Some cross-cultural work and developmental research has failed to find the diversity effect. Even with American college students, Osherson et al. (1990) reported an exception to the diversity effect: People draw stronger inferences given an observation that flies, for example, have some characteristic, compared to being given an observation that both flies and orangutans have this characteristic. It would be a challenge for any Bayesian account of induction, including Heit (1998), to address these exceptions, because Bayesian accounts do seem to predict robust diversity effects.

It is notable that Heit's (1998) Bayesian model of inductive reasoning predicts diversity effects without any size principle or assumption of strong sampling. Indeed, use of information about variability of evidence is taken to be a hallmark of Bayesian models in general (Earman 1992). Likewise, models of categorization without any size principle, such as Nosofsky's (1988b) exemplar model and Ashby and Gott's (1988) parametric model, also predict broader generalization from more variable observations. Although it is clear from **TENENBAUM & GRIFFITHS**'s Figure 2 that it is possible to predict the diversity effect with strong sampling and the size principle, it seems that the diversity effect in itself is not strong evidence for these assumptions. Other models without these assumptions can also predict this result.

The second, fascinating result, is that with other things being equal, larger samples tend to promote less broad inferences (reported in Tenenbaum 1999). This result does seem to be distinctive evidence for the size principle, as illustrated by **TENENBAUM & GRIFFITHS**'s Figure 3. This result would not be predicted by categorization models without the size principle such as Nosofsky (1988b) and Ashby and Gott (1988).

However, this result differs from numerous results showing just the opposite, with larger numbers of observations leading to broader generalizations. Although it is hard to perfectly eliminate confounds between number of observations and their variability, it appears that Homa et al. (1981) did show greater generalization to categories with more members. Nosofsky (1988b) showed that when a category member is presented a large number of times, there is increased generalization of similar stimuli to the same category. Maddox and Bohil (1998) showed that people can track the base rates of categories, with a bias to put transfer stimuli in more categories with more members. None of these results are insurmountable evidence against **TENENBAUM & GRIFFITHS**'s Bayesian model; for example, Bayesian models can easily incorporate information about base rates. Yet, it does appear that the result presented by **TENENBAUM & GRIFFITHS**, that larger samples lead to less broad generalization, is not characteristic of most results reported in this area. It would be important to establish the boundary conditions for this fascinating but isolated result.

In sum, the Bayesian model of generalization proposed by **TENENBAUM & GRIFFITHS** makes substantial contributions beyond existing accounts. However, the value of this model surely will be in its ability to address already documented phenomena in generalization, categorization, and inductive inference, including the exceptions to the diversity and sample size effects predicted by the model. In the target article, the model is applied to tasks where only positive cases of a single category are presented. Although it is valuable to focus on this important learning situation, it is notable that many more psychological experiments have addressed learning to distinguish one category from another, or learning from positive and negative examples. To address this large body of existing research, the Bayesian model itself would require some further generalization.

## Adaptation as genetic internalization

Adolf Heschl

*Konrad Lorenz Institute for Evolution and Cognition Research, University of Vienna, A-3422 Altenberg, Austria.* **Adolf.Heschl@kla.univie.ac.at**
**http://www.univie.ac.at/evolution/kli/**

**Abstract:** In the course of evolution organisms change both their morphology and their physiology in response to ever-changing environmental selection pressures. This process of adaptation leads to an "internalization," in the sense that external regularities are in some way "imitated" by the living system. Countless examples illustrate the usefulness of this metaphor. However, if we concentrate too much on Shepard's "universal regularities in the world," we run the risk of overlooking the many more fascinating evolutionary details which alone have made, and still make possible the evolution of diversity on earth.
**[SHEPARD]**

I will first attempt to theoretically underpin the concept of "internalisation" as it has been used and further developed in an impressive way by Roger **SHEPARD**. Let us begin at the lowest imaginable level of evolution: the genetic modification of organisms through random variation and natural selection. As has been empirically shown (Luria & Delbrück 1943), random genetic variation forms the molecular basis for subsequent evolutionary processes. These mutations can have three effects on the biological fitness of their carriers: they can be neutral, negative or – rarely – positive. In the first case, the traits of the organisms concerned will vary in a completely random way and no structuring effect of the environment will be recognizable. In the second case, many or even all carriers of the mutation will ultimately disappear from the evolutionary scene. This will have a clear structuring effect on the whole population of a given species, in the sense that only those individuals lacking this mutation will survive and reproduce. In such a case, we could speak of negative selection or, more simply, an extermination effect of specific adverse external influences.

The third case, in which a new mutation provides an advantage for the organisms concerned, leads to something we could indeed call an "internalization" of external regularities. To give an example at the molecular level: an enzyme (lactase) is produced which allows humans to better digest the form of sugar found in milk (lactose); as is to be expected, the distribution of the gene coding for this enzyme within the population reflects the structure of a concrete external regularity: the geographic distribution of intensive dairy-farming (Jones 1992). What is valid at the population level must also be valid at the molecular level: the chemical structure of lactase in turn reflects certain specific structural aspects of the disaccharide lactose. Hence, in an evolutionary perspective, it is perfectly legitimate to equate the process of biological adaptation with a kind of internalization process of external selection pressures, because *every* adaptive change must necessarily be accompanied by a corresponding form of internal restructuring.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Such a very general corroboration of **SHEPARD**'s main thesis on the evolutionary origin of certain perceptual-cognitive mechanisms even stands up to most of the both empirical and theoretical criticism advanced in the majority of the other target articles. This criticism basically concerns: (1) systematic misjudgements by psychological subjects in some experiments (cf. **HECHT**: gravity, horizontality, movement prolongation); and (2) the epistemological problem of how to provide sound empirical evidence for the existence of an internalization process (**HECHT, KUBOVY & EPSTEIN, SCHWARTZ, TODOROVIČ**). The first problem is not so severe, in the sense that one can easily imagine that every evolved mechanism, whether involved in the solution of physiological or behavioral problems, should also have been equipped during evolution with a set of necessary corrective procedures, which work well under certain circumstances, but not so well under others. Thus, we can assume that every internalized mechanism has a certain limited area of validity and if we leave this area or move towards its boundaries, we should not be surprised to see that we receive poor results (e.g., drawing a not perfectly horizontal waterline on a sheet of paper; Liben 1991).

The second, epistemological problem associated with the empirical support for internalization seems much more difficult, at least at first sight. It can nevertheless be resolved if we simply accept that which evolutionary theory tells us: that every existing perceptual-cognitive mechanism – learning and development included (Heschl 2001) – must have originated from an evolutionary process of adaptive internalization. The basic connection with evolution is readily explained: When we have detected an organismal trait (physiological, morphological, behavioral) which exhibits no random structural variation within a given population, we can assume that natural selection has shaped its specific form. Hence there is no way of demonstrating the internalization of a given trait (e.g., perception of apparent motion) by referring to another trait (e.g., kinematic geometry). In this sense **SHEPARD**'s thesis can indeed be neither confirmed nor refuted (cf. **HECHT**'s arguments). However, rather than searching for "absolute evidence" (what is after all absolute?) it is more feasible to investigate the manifold complex relationships between different cognitive mechanisms (cf. **KUBOVY & EPSTEIN**'s proposal to take kinematic geometry as a *model* of apparent motion). For example, there is little sense in opposing kinematic geometry to classical Newtonian physics, as does SHEPARD in his general statement that "geometry is more deeply internalized than physics" (what about our sense of gravitation?). It would be much more interesting to investigate *both* the conceptual connections *and* the behavioral transitions between them, if we assume that a basic modular structure (separate "areas") is combined with a hierarchical overall organization (connections) of the brain (Velichkovsky 1994).

Basically agreeing with **SHEPARD**'s internalization thesis, however, is not the same as accepting all his ideas about the evolution of perception and cognition. SHEPARD does not hesitate to speak of perceptual-cognitive "universals," and he concludes with the hope that one day psychology could even be investigated in a manner similar to elementary physics, finally achieving a comparable "mathematical elegance" and "generality of theories of that world." This hope, which is perfectly reductionistic and behaviorist at the same time, is at odds with evolutionary biology's stance concerning the fundamental causal units of biological evolution. It is not the universals or "natural kinds" (cf. Hull 1976) which drive development and change forward; individual genetic variation alone provides the raw material for natural selection. Hence, every new internalization must necessarily begin with a concrete individual organism which, by a genetic mutation, has been slightly altered with respect to the rest of the population. This means we are always confronted with (often unique) individual cases of internalized knowledge, from which we can of course try to distill the abstract case of a universally valid "type" (cf. Goethe 1795). It is nevertheless the individual cases which continue to constitute the real substrate of evolution.

To illustrate this it is sufficient to apply **SHEPARD**'s universal No. 3, the representation of objects "of the same basic kind," to the taxonomic group of Primates. Most members of this successful mammalian family can classify a large number of diverse inanimate and animate objects. But when we look more closely, we discover a series of additional differentiations. For instance, the ability to both discern and relate to one another an impressive number of different faces is one of the perceptual peculiarities of the human species, subserved by a specific anatomical region (Young & Yamane 1992). The same holds for humans' exceptional competence in learning to produce and recognize a variety of different single sounds (protophones, syllables) and sound combinations during early ontogeny, and the ability to apply them in symbolic communication (Oller 2000). On the other hand, the capacity to physically estimate very fine spatial relationships is concentrated only in certain primate species which are specialists in very rapid, yet secure tree climbing (marmosets, tamarins, spider monkeys, gibbons). Similarly, the most elaborate olfactory senses for distinguishing between tastes and flavours are found in the species which mainly eat fruit (Richard 1992). All this shows that there exists an incredible diversity of discernable internalized regularities which individuals of different species have successfully incorporated during their persistent "struggle for life." Universals are only one kind of abstract description of what really happens in nature. One could even say that the universalist approach must itself be the result of a specific phylogenetic internalization process during hominoid evolution. As such, it certainly fulfills its own adaptive functions.

## Group theory and geometric psychology

William C. Hoffman

*Institute for Topological Psychology, Tucson, AZ 85742-9074.*
**willhof@worldnet.att.net**
**http://www.home.att.net/~topoligicalpsychology/**

**Abstract:** The commentary is in general agreement with Roger Shepard's view of evolutionary internalization of certain procedural memories, but advocates the use of Lie groups to express the invariances of motion and color perception involved. For categorization, the dialectical pair is suggested.
**[BARLOW; HECHT; KUBOVY & EPSTEIN; SCHWARTZ; SHEPARD; TODOROVIČ]**

I whole-heartedly endorse **SHEPARD**'s view that over the long eons of evolution, certain psychologically significant universals have become internalized by veridicality. Your ancestor on the plains of Africa several million years ago, if indeed he/she *were* your ancestor, swiftly and appropriately adjusted to the image of a sabertooth tiger or gazelle and could tell red meat from green meat. Indeed, evolution is the basis for qualia. In the target article, I find much to agree with and yet some to quarrel with. In Stuart Anstis' celebrated phrase, "It *could* be that way, but is it really?"

Just as visual illusions are thought to reflect innate rather than veridical psychological structure, **SHEPARD** and his colleagues use apparent motion to provide a window revealing underlying perceptual mechanisms that normal perception masks. In particular, Chasles' law for kinematic trajectories, the Newtonian color circle, and psychological categorization ("kinds") are analyzed as products of evolutionary adjustment to the world's realities. Through it all runs the mathematical theory of groups (Shepard 1994, p. 12 ff.). "These are the group-theoretic principles governing rigid transformations in space, . . . what has been called kinematic geometry . . ." (Carlton & Shepard 1990a, p. 150).

**SHEPARD**'s use of anomalous perception as a window to the inner workings of the mind is in the best tradition of William James's view that: "To study the abnormal is the best way of understanding the normal." The practical implications for investigation of mental illness of the differences and commonalities between psychopathology and psychology, for example, could be tremendous.

The mathematical aspects of **SHEPARD**'s theory reach their apogee in Carlton and Shepard (1990a; 1990b). In the first of these


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

studies the history of group theory in psychology is reviewed (Carlton & Shepard 1990a, pp. 150–51, 177). My early work (Hoffman 1966) on the Lie transformation group theory of perceptual neuropsychology is dismissed as too neuropsychological, the focus being instead on descriptive *psychological* "system" without mechanism. Yet perception revolves around psychological constancy, and the Neuron Doctrine is fundamental to psychological science. The neuropsychological system is par excellence an integrator of local (neuronal) to global (psychological). Not only are Lie groups the classic mathematics for relating local to global – neuronal to psychological – but they also constitute the natural mathematical structure for expressing *invariance:* "universal," "regularity," "object constancy," "color constancy," "symmetry," "conservation law," and so on. (Hoffman 1998). The group's Lie derivative *annuls* a perceptual invariant just as the symmetric difference nulls a cognition matched with memory (Hoffman 1997; 1999). And the Lie transformation group appropriate to perceptual invariance – psychological constancy – is simply the conformal group CO(1,3) (Hoffman 1989; 1994). Motion perception – and object invariance under motion – are governed by a perceptual Lorentz transformation (Caelli et al. 1978a), and in this mileu a moving object *can* "travel too fast for the eye to follow," owing to the presence of the Lorentz factor. Conformality is known to be equivalent to causality (Guggenheimer 1977, p. 247:3) and also admits **KUBOVY & EPSTEIN**'s E → S → E. Curvature, given by the rate of rotation of the tangent to a visual contour, is a direct consequence of motion constancy and cortical orientation response fields, and is the simplest metrical invariant of a curve (Gamkrelidze 1991, p. 26). A Pleistocene hunter could not rely on fitting circles to his spear's trajectory but had to take account of the actual curved path, thus internalizing over generations of those who survived the procedural memory for the motor skill of throwing (Calvin 1990). Chasles' law provides what is called in economics a turnpike theorem, wherein the turnpike (kinematic "geodesic") undergoes a small correction at the end.

**SHEPARD**'s second universal is color constancy, given by the quotient group SO(3) / S(O) over the Newton color cone. Now the Benham top, which generates *color* by rotation of *black and white* patterns, and Daw's finding that certain cells in the visual pathway respond to both motion *and* color suggest that a hyperbolic geometry, like that of the subjective Minkowski space for motion perception, is more appropriate than Euclidean. The Newtonian color circle should therefore be replaced by a Poincare disk. The gap of extraspectral purples in the Newtonian color circle then disappears and opponent pure colors can be directly joined, as opponent color theory requires, by hyperbolic "lines" (circular arcs in the Poincare disk). And, in accord with the color parameters of hue, saturation, and brightness, the hyperbolic "lines" do pass through the neutral gray center of the hue circle.

**SHEPARD**'s third universal is that of an "object's kind," its categorization. Hoffman (1999), argues that psychological categories and mathematical categories (the most general sorts of equivalences) are essentially the same. Both have objects (things) and morphisms (mappings), but in mathematical categories the objects at the vertices of a connectionist diagram are secondary to the arrow-structure, just as in the cognitive developmental sequence, where childhood "eidetic imagery" (objects) fades as we age – "forgetfulness" – in favor of enhanced trains of thought (the mappings). **SHEPARD**'s characterization of objects of the same kind (Shepard 1994, p. 22) as those providing the *same function,* not necessarily the same appearance nor the same attributes, fits well. **KUBOVY & EPSTEIN**, too, note "that an evolutionary stance helps focus attention on function." Basic level categories are further characterized (Shepard 1994, p. 23) in terms of "a dichotomous connectedness," which is in accord with the dialectical-pair model for cognition (Hoffman 1999) but without need for metrics or multidimensional scaling. Finding the ponderous multivariate calculations of MDS internalized in actual brain tissue would be surprising indeed. As **TODOROVIČ** (this issue) observes, ". . . what would impel a biological visual system to pose, let alone solve such a problem?" However, the simplicial fibrations (Rusin 2000) that guide whatever thought processes *are* involved certainly appear present in so-called "brain circuits." Once more the constraints imposed by neuropsychology help to mark the way.

Two reservations with regard to **BARLOW**'s article: "Statistical regularities of the environment" and his view of movement perception as resulting from cortical computation of the difference quotient. As Boring (1942, p. 595) notes, Wertheimer's finding of apparent movement shows that "movement is movement." Perception of successive positions is not essential, for then *apparent* movement could not occur. Movement perception, embodied as the Lorentz group (Hoffman 1978) in the small pyramid cytoarchitecture of the visual cortex, is intrinsic. Excitation of this cytoarchitecture at an appropriate rate generates path-curves characteristic of apparent motion (Caelli et al. 1978b). Frequency of neuronal firing codes psychophysical intensity, not shape. For that you need the integration of orientation response fields into visual contours by parallel transport with respect to the constancy-generated vector field. As to *statistical regularity,* perceptions occur in real neurobiological time, without the time-consuming computations required for estimation of statistical parameters. The same applies to **TENENBAUM & GRIFFITHS'** Bayesian priors. I see no way in which statistical laws – **HECHT**'s "ill-resolved statistical regularities" – extending over the eons involved in evolution could emerge in the face of marked individual variation and widely varying cultures. Evolutionary adaptation to regularities, yes, but not *statistical* regularities without convincing proof that subjective statistical decision theory was internalized in Pleistocene hunter-gatherers. It is rather more likely that in real neurobiological time the brain acts to smooth the commonality of short term memory (STM) and working memory (WM).

**HECHT** interprets "internalization" as habit but neglects a more likely candidate, namely, instinct – territorial, maternal, body language, and above all in the present context, the survival instinct. He claims that no evidence exists for innate visual responses to gravity and horizontality. Yet, numerous experiments by psychologists have confirmed that gravity and vision interact. Indeed the circular path that was chosen by "many subjects" for a ball emerging from HECHT's C-shaped tube makes sense if the tube is vertical and gravity is acting. Just as nearly half the population never achieves Piagetian formal operations, so too Hecht's water level task was solved by only half his subjects. Yet the experimental protocol appears to be such as to confuse visual illusion with knowledge of physical law. It would be interesting to perform this experiment in the context of Witkin's field dependence (Witkin & Goodenough 1981). Despite Hecht's impression that he has "attempted to focus on natural viewing situations," artificial laboratory contrivances and natural contexts (Neisser 1982) are once again confused. Hecht's "externalization of body mechanics" as an alternative to **SHEPARD**'s "internalization" seems simply a play on words. Both HECHT and SHEPARD should find Adair (1990) interesting.

The philosophical bent of the **KUBOVY & EPSTEIN** article offered considerable difficulty. Some of the ideas are even mutually contradictory, for example, the lack of an inverse to E → S even in the presence of recognition; the view that perceptual laws and the intellectual activity of their discovery are unrelated; and the conception that kinematic geometry is a superset of the perception of real motion rather than being a subset of dynamics of a rigid body (Banach 1951, Chapters VII and VIII; and **TODOROVIČ**, this issue). Both real and apparent motion perception are simply different mappings from the cortical embodiment in the progression E → Retina V1 → V2 → V3/VP → MT → V4 → S of the subjective Lorentz group. KUBOVY & EPSTEIN suggest that the term "internalization" may lead to confusion and they use the metaphor of emotion, which unfortunately *is* evolutionarily internalized (Damasio 1999, pp. 54–59). Comparative psychology and comparative neuroanatomy (MacLean's triune brain) have a long and honorable history. APA Division 6 is devoted to precisely this area. Artistic skill is internalized, beginning with Pleistocene Lascaux cave art and evolving historically in both skill and imagination. The fact that "the confusion has not abated" with respect to mental ro-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tation is surely significant. Kubovy & Epstein are critical of the metaphor they sense in **SHEPARD**'s approach and suggest that science progresses through the interaction of better data and new theory, yet there seems to be a good deal of philosophizing and play on words in their own article (see Synge 1951). In my view, *cognitive* psycholinguistics has little to do with the *perception* of motion. Hence their attempt to assess what guides SHEPARD's thinking seems to me irrelevant. Their *Metaphors of mind* seems to me another play on words, at odds with their own recommendation to strip scientific writing of metaphors.

Robert **SCHWARTZ**'s article is tightly reasoned and insightful. I suggest that his "universal regularities of the external world" represent, for perception, the distortions imposed by viewing conditions which the psychological constancies correct – "under normal viewing conditions, a real object that deforms its shape will generally be perceived as such." Dialectic plays a similar role (Hoffman 1999) for the universality of uncertainty and contradiction. Assimilation and accommodation play an integral role in cognition, and veridical adjustment to the actual geometrico-physical world that we live in does confer an evolutionary advantage. Your ancestor did not walk off a cliff or miss his spear-throw and get trampled by the mammoth. SCHWARTZ's interpretation of **SHEPARD**'s kinematic principle as a turnpike theorem is valid. But even in the perception of actual motion, there are perceptual anomalies, for example, the subjective Fitzgerald contraction and time dilation of fast moving objects (Caelli et al. 1978a). Schwartz's view that models of visual processing and underlying mechanisms can be formulated and tested independently of issues of origin is correct, but *empirical import* and *systematic import* (Hempel 1966) are thereby lessened.

**TODOROVIČ**'s article is a model of precise thought and analysis, both philosophical and mathematical. His view that some perceptual competencies are internalizations of external regularities as invariances extending over evolutionary history is well based on paleoanthropology (Fomenti 2000). Certain aspects of "circular translation" suggest the "cranks" of the configuration spaces of linkages (Thurston & Weeks 1984), and translations + rotations are affine transformations (Eisenhart 1961, p. 43:7). More realistic psychologically than rotation + translation may be the dilation group for size constancy + the rotation component of the Lorentz group. The resulting orbits are the spirals universal in nature, not the least of which may be Grinvald's "pinwheels" of cortical orientation response fields. I like TODOROVIČ's clear discussion of kinematic geometry versus classical physics but would add something on topology, surfaces, and groups (cf. vol. 3 of Aleksandrov et al. 1969). Todorovič's general idea that, through contact with the environment, our perceptual systems have extracted certain principles that are strongly analogous to laws, axioms, and theorems articulated by the scientific community, is reminiscent of Shulman's three kinds of knowledge: propositional, case, and strategic. Propositional knowledge is essentially deductive and consists of principles, maxims, and norms. Case knowledge is commonsense extrapolation of detected regularities via analogical reasoning to representative rules. Strategic knowledge is a dialectical metaknowledge that comes into play when principles and case rules collide or contradict.

## Learning to internalize: A developmental perspective

Bruce Hood

*Department of Experimental Psychology, University of Bristol, Bristol, BS8 1TN, United Kingdom.* **bruce.hood@bristol.ac.uk**
**http://www.psychology.psy.bris.ac.uk/psybris/BruceHood.html/**

**Abstract:** As Hecht points out, finding unequivocal evidence for phylogenetic knowledge structures is problematic, if not impossible. But if phylogeny could be dropped, then internalization starts to resemble the "theory theory" approaches of developmental psychology. For example, an appreciation of falling objects leads to a very strong bias that could be regarded as internalized knowledge acquired during ontogeny.
**[HECHT; SHEPARD]**

In support of his argument against **SHEPARD**'s notion of internalization, **HECHT** has used the three examples of gravitational acceleration, appreciation of horizontality of liquids, and predicting trajectories of objects in motion. These have been selected as test cases as they reflect universal invariant properties of the real world, and hence, should be ideal candidates for internalization. However, in each of these three different examples, individuals fail to perceive, and in the latter two cases, to predict the correct invariant laws. HECHT concludes that the three examples show that the internalization hypothesis is in trouble. But are these examples sufficient to counter the notion of internalization entirely? If internalization reflects evolutionarily contrived solutions, then there must be good reasons to assume that an appreciation of these invariant laws would have been worth internalizing. Also, in the third example of object motion, there does appear to be an appreciation of Newton's laws of motion if individuals can observe possible and impossible trajectories rather than making predictions.

Aside from these test cases, **HECHT** is correct in arguing that **SHEPARD**'s notion of internalization is difficult to falsify. As Hecht points out, if one did find a candidate for internalization, it would have to be demonstrated that it would not have been internalized by default. To quote Hecht, "the organism must have had a chance to fail to internalize the knowledge in question." Furthermore, it would have to be shown that it was acquired in the absence of other learning processes. These qualifiers seem to be strong barriers against ever finding support for internalization, as it is difficult to imagine the evidence that would ever satisfy these criteria.

However, if we loosen the criterion and allow the internalizing process to emerge within ontogeny, then we may still have a useful construct, though it would appear that such a proposal may be anathema to **SHEPARD**'s concept of internalization. In the field of cognitive development, there are a number of knowledge structures that serve to identify environmental regularities and constraints and help to solve the underspecification problem. Instead of internalization, developmentalists use the term "theory theory" to describe, in very simple terms, the process whereby normal children come to internalize knowledge structures that help to organize and interpret the world. This is regarded as internalization rather than simple learning, or Jamesian habit formation, because one criterion of "theory theory" is that once acquired, these structures are difficult to adapt – a feature that **HECHT** offers as a sign of internalization. Furthermore, as these knowledge structures are acquired within the lifespan of the individual – and not invariably for all, – the potential to fail internalization is always present.

For example, an appreciation of gravity may not be inbuilt. Infants below 7–8 months of age do not reliably detect that unsupported bodies must fall (Spelke et al. 1992). Even Piaget (1954) first reported that prior to this age, infants do not anticipate that released objects must fall down. However, soon afterwards, possibly as a result of extensive experience of dropping objects themselves, children start to form a very strong bias towards assuming that all dropped objects fall in a straight line. There is also some evidence for a similar but weaker bias for objects travelling horizontally (Frye et al. 1985; Hood et al. 2000).

By 2 to 3 years of age, if an object is dropped into a curved conduit such as a tube (see Fig. 1), children invariably search for fallen objects straight down even when they are given numerous trials and experience with the tube (Hood 1995). If the motion of the object is reversed so that object appears to be sucked up the tube, the bias disappears (Hood 1998). This is clear evidence for theory-like reasoning about the trajectory for falling objects. While children eventually solve the tube task by approximately 3 years of age, the internalized structure is still present, as additional task loads in older children (3–5 yrs; Hood 1995) reveal that the gravity or straight-down bias is still operating. The same straight-down bias probably underlies the cliff studies in older children (4–12 yr-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press



Figure 1 (Hood). Child searching for a fallen object on the Tubes Task.

olds; Kaiser et al. 1985b) and in adults who have been asked to predict the trajectory of objects dropped from a plane, though there have been alternative accounts based on perceptual illusion (McCloskey et al. 1983).

The straight-down bias for falling objects is a candidate for internalization. It is invariant and shared by at least two different species (Hood et al. 1999). It is a useful rule of thumb for predicting object locations if the trajectory is underspecified. It remains resistant to counter-evidence and probably continues to operate throughout the individual's lifespan but is not necessarily built-in. And there are other domains of knowledge that have the same profile as falling objects. In developmental psychology, these are regarded as internalized knowledge structures. But if one sticks to the strict criterion of phylogenetically specified mechanisms, it seems unlikely that **SHEPARD**'s internalization will ever be falsifiable. While an appreciation of falling objects may fit with **HECHT**'s externalization hypothesis, most "theory theory" phenomena in cognitive development do not readily lend themselves to an analysis based on motor control, and so may not help with Hecht's alternative approach.

ACKNOWLEDGMENT

This research has been supported by the NICHD and a Sloan Fellowship.

## Internalized constraints in the representation of spatial layout

Helene Intraub

*Department of Psychology, University of Delaware, Newark, DE 19716.*
**intraub@udel.edu**

**Abstract:** Shepard's (1994) choice of kinematic geometry to support his theory is questioned by Todorovic, Schwartz, and Hecht. His theoretical framework, however, can be applied to another domain that may be less susceptible to some of their concerns. The domain is the representation of spatial layout.
**[HECHT; SCHWARTZ; SHEPARD; TODOROVIČ]**

In their insightful discussions of **SHEPARD**'s (1994) theory, **TODOROVIČ, SCHWARTZ,** and **HECHT** raise different but interrelated concerns about **SHEPARD**'s choice of kinematic geometry to support his ideas about the internalization of external regularities. These include questions about the "uniqueness" of motion in the world, the ecological validity of apparent motion, and whether the theory can be falsified. There is controversy as to whether examples from motion perception are as strong as his example of the sleep-wake cycle. **HECHT** calls for other domains to be specified. I propose that representation of spatial layout may be a good domain to consider in this context.

***Perceiving spatial layout.*** The world is continuous and is packed with detail, but our view of the world is not. We cannot perceive our surrounding environment all at once. In vision, ballistic eye movements shift fixation as rapidly as 3–4 times/second. Even while fixating a specific location, high acuity is limited to the tiny foveal region (2 degrees of visual angle) and drops dramatically outside the fovea, yielding a large low-acuity periphery. In haptic exploration (without vision), hands can touch only small regions of the surrounding world at a time. Yet, these successive inputs support a coherent mental representation of the surfaces and objects that make up natural scenes. Research on transsaccadic memory and change blindness (e.g., Irwin 1991; 1993; O'Regan 1992; Simons & Levin 1997) supports the idea that mental representation is more schematic and abstract than perceivers realize (e.g. Hochberg 1986). I propose that in addition to maintaining layout and landmarks from prior views of the environment, mental representation includes anticipatory projections about future views. These projections are internally generated but are constrained.

Evidence for this is provided by a common representational error that occurs in memory for photographs of scenes, referred to as "boundary extension" (Intraub & Richardson 1989). Viewers remember having seen a greater expanse of a scene than was shown in a photograph. What is important to note is that different viewers all seem to make the same unidirectional error. For example, in one of their experiments, out of 133 drawings made by 37 individuals at least 95% included this unidirectional spatial extrapolation (examples are shown in Fig. 1). In recognition tests viewers tend to rate the same view as showing too little of the scene, and will frequently select a more wide-angle view as the one they saw before (e.g., Intraub et al. 1992). Boundary extension occurs following brief (e.g., 250 msec) and long (e.g., 15 sec) presentations and occurs as rapidly as 1 second following picture offset (Intraub et al. 1996). The adaptive value of this surprising distortion is that although the mental representation is inaccurate with respect to the photograph, it contains a remarkably good prediction about the scene that the camera partially recorded.

The observer's representation is constrained to shift outward rather than inward. This may reflect the internalization of the spatial continuity inherent in our environment. It is a universal regularity of the world that there is always more just beyond the current view. Anticipating layout would facilitate integration of successive views, and would help draw attention to unexpected features that arise when the anticipated region is actually scanned. A perceptual system with a small, highly focused sensory area that actively explores the world would function with greater economy if the border of the current view were ignored. Indeed, boundary extension does not appear to be cognitively penetrable. Even with forewarning and prior experience, viewers were unable to prevent its occurrence (Intraub & Bodamer 1993).

***The "uniqueness" of spatial continuity.*** In his critique of the notion of "kinematic uniqueness," **TODOROVIČ** questions whether one can determine a priori which type of motion would become the one to be internalized. He describes various types of motion and asks why the perceptual system would prefer one type rather than the other. Spatial layout provides a regularity of the world that seems unequivocal. It is so fundamental that it sounds trivial; wherever one looks in the environment there is always more. Whether one is in an enclosed space or an open field, small changes in the position of the head or eyes will bring a new region into view. The continuity of spatial layout in the environment



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

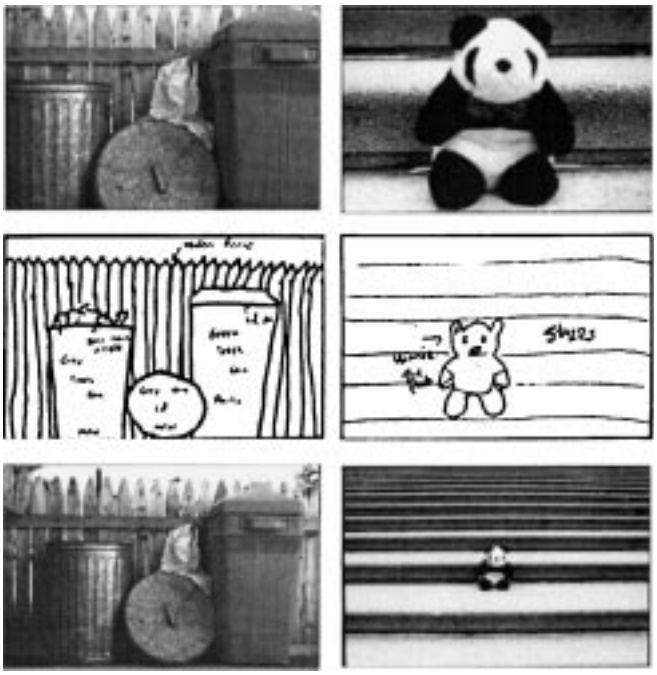


Figure 1 (Intraub). Top row shows stimuli, middle row shows representative subjects' drawings from memory of those stimuli, bottom row shows a more wide-angle view of the scene. Note that the remembered pictures contain information that actually did exist outside the borders of the original view. Column 1 pictures are from Intraub and Richardson (1989) in which there were 15 second stimulus durations, and Column 2 pictures are from Intraub et al. (1996) in which there were 250 msec stimulus durations.

seems to provide a less debatable starting point for considering a regularity of the world that might be a good candidate for internalization.

***Ecological validity.*** Among his concerns, **SCHWARTZ** questions the ecological validity of **SHEPARD**'s focus on apparent motion. He is unconvinced that the same constraints seen in apparent motion are necessarily implemented during more normal instances of motion perception in a rich well-illuminated environment. Yet, the paradigm requires the removal of information in order for the constraints to be seen. Can internalized constraints be tested without artificial circumstances? In the case of boundary extension, in one sense, looking at a photograph is like looking at the world through a window (with the borders occluding all but the exposed area). What is useful about the paradigm is that we can test the internalization hypothesis under normal viewing conditions in the three-dimensional (3-D) world. The basic question is whether, under conditions that allow for stereopsis, motion parallax, and the ability to gauge sizes and positions with respect to one's body, would the viewer experience boundary extension?

To answer this question, viewers studied bounded regions of six real scenes made up of common objects on natural backgrounds. An occluding window was placed around each scene, thus exposing an area and occluding the surrounding space. The subjects directly looked down at the scenes, which were arranged either on tabletops or the floor. After they studied the scenes, the occluding windows were removed. Viewers returned to the same position in front of each scene and indicated how great an expanse they had seen minutes earlier. Occluding borders were placed at the designated locations and the viewers made any adjustments necessary to ensure that the exposed space was the same they remembered seeing minutes before. Subjects clearly remembered having seen a greater expanse of the scenes than they actually had – increasing the exposed area by 45% (see Intraub, in press).

***Generalizing across sensory modalities.*** If spatial continuity is a unique regularity of the world that is an internalized aspect of representation, then we would expect to see it underlie perception of layout in any modality suited for detecting layout – that is, vision and haptics. To determine if haptic representation would show evidence of anticipatory spatial representation, we conducted the same experiment described above with an individual who has been deaf and blind since early life, as a result of a genetic disorder (Leber's Syndrome). A control group of blindfolded-sighted subjects also participated. Would someone whose experience with layout is haptic show the same error experienced by sighted subjects? She must also integrate successive inputs as her hands explore spatial layout. And the external regularity of continuity is the same irrespective of modality. What happened is that both she and the control subjects remembered having touched a greater expanse of the scene than they actually had. In this case the control group increased the area of the exposed region by 32%. The deaf/blind observer's representation of the exposed regions was remarkably similar to the control subjects'. She matched or exceeded their boundary extension on all but one scene (see Intraub, in press).

***Can the internalization hypothesis be falsified?*** **HECHT** questions whether the internalization hypothesis can be falsified or not (i.e., Popper's test). He argues that the resolution with which one can specify an internalized regularity will determine to what extent it can be experimentally tested. In the case of layout, we cannot specify a metric that will predict exactly how much extrapolation will occur. However, we can make predictions about patterns of responses under conditions in which the internalized regularity applies, and when it does not. In other words, we can articulate a "boundary condition" for this outward extrapolation. Drawings of objects on scene backgrounds (i.e., backgrounds that depict part of a continuous location) should give rise to boundary extension. In contrast, drawings of the same objects on blank backgrounds (object not in a depicted location) should not. If boundary extension occurred in memory for pictures in which no location was depicted, then there would be no principled argument to support the contention that boundary extension reflects an internalized constraint about the world.

Intraub et al. (1998) conducted a series of experiments to test this contrast. Boundary extension occurred in memory for pictures with the surfaces depicted in the background, but not for those with blank backgrounds. In the latter condition a unidirectional distortion was not obtained, instead, size averaging occurred (larger objects were remembered as smaller and smaller objects were remembered as larger). In other experiments in the series, we tested Shepard's (1984) proposal that imagination should draw upon the same internalized constraints as does perception. We found a striking effect of imagination instructions in this task. Subjects viewed the same drawings of objects on blank backgrounds; one half imagined natural backgrounds behind the objects while they viewed them, the other half imagined the objects' colors while they viewed them. Imagining scene background resulted in boundary extension, whereas imagining object colors resulted in size averaging. Mental representation of the same objects was affected in a predictable manner depending on whether or not the subject imagined layout.

In conclusion, perception of spatial layout appears to be a plausible domain for testing ideas about internalized constraints. It provides a complementary approach to the problem. Instead of depriving the viewer of information so that we can reveal the underlying constraint, we examine the constraint through evaluation of a normally occurring "error."

ACKNOWLEDGMENT

Preparation of this commentary was supported by NIMH MH54688-02.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

# Reliance on constraints means detection of information

David M. Jacobs,[a] Sverker Runeson,[b] and Isabell E. K. Andersson[b]

[a]*Vrije Universiteit, Amsterdam, Faculty of Human Movement Sciences, 1081 BT Amsterdam, The Netherlands;* [b]*Uppsala University, Department of Psychology, SE-75142 Uppsala, Sweden.* **d_jacobs@fbw.vu.nl {sverker.runeson; isabell.andersson}@psyk.uu.sewww. http://www.fbw.vu.nlwww. http://www.psyk.uu.se**

**Abstract:** We argue four points. First, perception always relies on environmental constraints, not only in special cases. Second, constraints are taken advantage of by detecting information granted by the constraints rather than by internalizing them. Third, apparent motion phenomena reveal reliance on constraints that are irrelevant in everyday perception. Fourth, constraints are selected through individual learning as well as evolution. The "perceptual-concept-of-velocity" phenomenon is featured as a relevant case.

**[HECHT; KUBOVY & EPSTEIN; SHEPARD]**

**SHEPARD** proposes internalized constraints as the vehicles by which perception achieves concordance with the environment. His prime examples are apparent motion phenomena in which participants perceive motion paths that, some would say, are not present in the stimuli. As pointed out by **SHEPARD** (p. 585, target article), and by **KUBOVY & EPSTEIN** (**K&E**) as well, these paths are noticeable only as visual defaults in severely impoverished vision conditions.

A motion perception phenomenon that occurs under normal conditions, yet entails large discrepancies with the physical motion, is the so-called "natural start" or "perceptual concept of velocity" phenomenon (Runeson 1974; 1975). Motions that start in an inertially natural way, that is, with a constant acceleration that gradually levels off to a constant velocity, appear to have constant velocity throughout. Motions with other velocity profiles appear accelerated or decelerated depending on how they differ from natural start motions.

In both apparent motion and the natural start phenomena, perception seems to "go beyond the given" in an adaptive way. By this, a role for environmental regularities in perception is indicated. We argue, however, that the phenomena do not prove internalization of these regularities.

***Reliance on constraints: Indispensable or exceptional?*** The biological purpose of perception is to establish informational coupling between organisms and their environments. *Information* is the commodity perception deals in. Regularities, or constraints, are the "grantors of information" (Runeson 1988; 1989).[1] If an organism uses a particular informational variable, it demonstrates dependence on the prevalence of the constraints that grant the usefulness of the variable. We do not disagree, therefore, with the claims in the target articles that perception takes advantage of constraints. To the contrary, even stronger claims should often be made. Constraints are necessities (e.g., Runeson 1988; 1994; Runeson et al. 2001). Since there can be no information without constraints, there can be no perception without constraints. Two examples illustrate that this stronger claim is not always appreciated.

First, reliance on constraints is often considered necessary only when the stimulus in itself does not specify the to-be-perceived property. For instance, in describing **SHEPARD**'s position **HECHT** writes, "Whenever the stimulus is ambiguous or ill-defined, as in apparent motion, an internalized default influences the percept" (p. 609). Such claims seem to imply that some stimuli can be unambiguous by themselves, that they can be specific to properties in a supposedly unconstrained world. This, however, is not the case.

Second, reliance on constraints is also invoked as an explanation for perceptual deviations from the "objective stimulus." Proponents of different theoretical views would agree, we think, that the natural start phenomenon shows reliance on the constraint that material motions must start gradually. Imagine that the empirical situation had been the opposite, that is, that starting motions had been perceived in accordance with the physical-science concepts of velocity and acceleration. Would it be agreed that constraints are required to explain that too? Presumably not, but why not?

Runeson (1974; 1975) proposed that motion perception is couched in terms of a *perceptual concept of velocity* (PCV), which incorporates the characteristic way in which motions of material objects start. The PCV has superior descriptive power because the speediness of most pre-technological motions is describable with a single constant value. However, descriptions couched in physics-velocity terms are commonly considered real and objective, whereas PCV-based descriptions seem derived or subjective. The PCV phenomena thus may appear to require inferential conversion, which would necessitate constraints, while physics-velocity conformant percepts would not.

Such a distinction is not valid, however. Scientific concepts are invented and defined to provide convenient ways of handling various phenomena under study, hence they are on equal ontological footing. Whether or not additional processing is needed for physics-velocity- or PCV-style perception is not deducible from basic physical laws but depends on which variable the measuring device is designed/developed for (Runeson 1977; 1994).

In sum, all perception depends on constraints – without regularities, stimuli are always ambiguous. Accordingly, as we have discussed elsewhere (Runeson 1988; Runeson et al. 2001; see also **K&E**), a cornerstone in Gibson's groundbreaking contributions to perceptual theory was the invocation of appropriate constraints – although not done in current terminology (see Gibson 1950; 1966; 1979). One can debate what regularities are important and how they constrain perception, but not whether or not perception relies on environmental regularities.

Relatedly, **HECHT** draws attention to experiments in which perceivers appeared not to rely on a few presumably useful constraints and argues that this is problematic for internalization theory. We do not agree. Any theory that poses reliance on constraints should accommodate the fact that not all constraints are relied on. Perception always takes advantage of constraints, but this does not imply that *all* conceivable constraints are taken advantage of.

***How to take advantage of constraints?*** Theories that describe how constraints are taken advantage of can crudely be classified as either "internal-entities theories," which hold that perception draws on knowledge or assumptions *about* the constraints, or "regularities-in-design theories," which hold that the design of organisms is compatible with the constraints. Both types of theories have their merits. Internal-entities theories are more useful if one considers cognitive processes. A physicist, for instance, can use knowledge of laws to make inferences about, say, the amount of kinetic energy of a system. The regularities-in-design alternative is better suited to describe, for instance, how lungs take advantage of the availability of oxygen in the air – without knowledge about that constraint.

Which analogy is better for perception? **SHEPARD**'s internalization notion seems to be an internal-entities theory. However, in keeping with the ecological approach, and with **K&E**, we are skeptical about internal entities for the explanation of perception. A long explanatory story would have to be told, going from constraints, to informative variables, to organs that can register the variables. We see no way that such internalized constraints could be beneficial for the perceiver and, therefore, no way that evolution could have endowed us with them. Evolutionary changes in the constitution of visual systems must be on account of adaptive advantages gained from modifications in the pickup of optical variables and how these are used in action control. Thus, the evolution of perception is not a matter of constraints getting internalised. Rather, it is the ways of exploiting the external *consequences* of constraints – specificity of optical variables – that develop and thus upgrade perceptual skills.

The regularities-in-design alternative is to some extent implicit also in most internal-entities theories; realistically, there would be too many constraints for a perceiver to consider explicitly (Runeson 1988). This is evident, for instance, in the lens structure of the



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

eye, which takes advantage of the refraction properties of light, without knowledge about these properties. Thus, an additional burden for internal-entities theories is that they should explain why perceptual systems take advantage of some constraints by carrying assumptions about them, and of other constraints without such assumptions.

Our understanding of the role of constraints entails a different idea about the division of labor between perceivers and perception scientists (likewise, see **K&E**, p. 618, Abstract). Perceivers are just being selected or reinforced for the information they use. It is we, the scientists, who need to bring in constraints to explain perception – just as in explaining any other real-world phenomena.

***The importance of apparent motion.*** The previous section argued that perceptual systems take advantage of constraints merely by relying on the information granted by them. We now consider implications of this view for **SHEPARD**'s interpretation of apparent motion phenomena. SHEPARD claims that apparent motion results reveal deeply internalized universal principles. Our view, in contrast, is that they reveal a reliance on constraints that are usually not relevant in more natural situations.

Many potentially useful constraints exist. The target articles consider, for instance, object constancy, theorems of kinematic geometry, gravitational acceleration, and so on. Previously considered constraints include the rectilinear propagation of light, conservation of momentum, and the regularity of surface texture. Such lists could be extended endlessly. Given the diversity of constraints, one might wonder whether all types of constraints are equally important for perception. Can constraints be classified in a way that enhances our understanding?

Classification can, for instance, be based on the extent to which there are exceptions to constraints (e.g., **HECHT**). If there are exceptions, then the constraint is not sufficient to grant full specificity of information. The general veridicality of everyday perception seems to show that perceivers use information granted by constraints which apply to a large extent in the relevant ecologies.

This is not the case, however, if the typical richness of information is artificially reduced. What happens if a visual system is deprived of its usual optical support? Does it give up and is nothing perceived? No. Pressed to function, the visual system reverts to other variables. Since few variables are available in the impoverished situation, the variables used might not be "information" in the specificational sense.

In apparent motion stimuli, typically used information is not available. Therefore, the perceptual system is forced to use other optical variables, which rely on other constraints. These might include theorems of kinematic geometry, as argued by **SHEPARD**. But, even if this were the case, we argue that these constraints are not relevant in regular motion perception. Participants are merely forced to rely on such constraints by the impoverished stimuli. It follows that apparent motion phenomena do not necessarily reveal fundamental characteristics of vision – the opposite might very well be the case.

***Universal principles of mind?*** One of **SHEPARD**'s main goals is to reveal universal principles of mind. He argues that such principles can exist because many constraints are universal and that, consequently, evolution might have shaped the minds of all species to reflect the same universal constraints. In opposition, we argue that many more constraints prevail in the niches of particular species than in the universe. For an animal and its evolution it does not matter whether or not a constraint applies outside the niche. One could expect that animals often rely on local constraints that apply only in their niches or, in other words, that animals often use variables that are useful only there. Furthermore, perceivers can learn to take advantage of the particular constraints in different task situations (e.g., Jacobs et al. 2000; Michaels & de Vries 1998; Runeson et al. 2000).

In sum, constraints that are relied on depend on the particular ecology in which the species evolved as well as on the learning history of the individual. This indicates that the minds of individuals are just as likely to reflect local, as universal, constraints – a discouraging perspective if one searches for **SHEPARD**'s type of mental universals.

On a positive note, we suggest that although individuals might differ in the constraints they exploit, universal principles of mind might reside elsewhere. For instance, some principles of learning might hold very widely. One of these could be that the looser the exploited constraint, the faster perceivers learn not to rely on it. Or equivalently, the poorer the detected variable, the faster perceivers come to detect other, in the long run better, variables. We suspect that searching for principles at this level is more fruitful than trying to determine which regularities are reflected in the mind, the way suggested by **SHEPARD**.

NOTE

**1.** Kubovy & Epstein use the word "guarantors." We consider the improvement subtle and retain the original term "grantors of information" (Runeson 1988).

ACKNOWLEDGMENTS

The writing of this commentary was supported by grants from The Netherlands Organization for Scientific Research (NWO grant no. 575-12-070), and from the Swedish Council for Research in the Humanities and Social Sciences (HSFR). We also thank Claire Michaels and Rob Withagen for helpful comments on a previous version.

## What is internalized?

Mary K. Kaiser

*NASA Ames Research Center, Moffett Field, CA 94035.*
**mkaiser@mail.arc.nasa.gov**
**http://www.vision.arc.nasa.gov/~sweet/kaiser.html/**

**Abstract:** Hecht provides insights concerning the difficulty of empirically testing Shepard's internalization hypothesis, but his argument for an externalization hypothesis suffers from similar sins.
**[HECHT]**

**HECHT**'s article does a fine job of focusing the reader on the two primary questions raised by **SHEPARD**'s internalization hypothesis. These are: (1) What is "internalized"? and (2) *What* is internalized? As for the definition of internalization, HECHT contrasts literal and abstract interpretations, but acknowledges that these are actually two ends of a continuous spectrum. Nonetheless, once one deviates from a literal interpretation, the hypothesis becomes more difficult to falsify.

This raises the question of whether falsification is as crucial a criterion for theoretical utility as **HECHT** assumes. The value of a psychological theory rests in its ability to describe and predict behavior. Yes, a theory must be testable, but ultimately, all that can ever be tested is a particular instantiation of a theory. The failure of these instantiations undermines the theory's utility as a descriptive and predictive device.

In fact, no instantiation of the internalization hypothesis has proven robust to empirical testing. However, this reflects the flaw of the particular instantiation, not the underlying hypothesis. **HECHT** argues that the logical opposite of internalization is externalization (i.e., the observer's imposition of his own body dynamics on an under-specified stimulus to create the perceived reality). Actually, the logical opposite of internalization (as well as externalization) is an unconstrained perceptual system – one that imposes no assumptions, and finds any under-specified stimuli ambiguous and uninterpretable.

Obviously, our visual system is not unconstrained. It both filters incoming stimuli (e.g., extracting zero- and first-order kinematics while virtually ignoring higher-order motions), and imposes assumptions, biases, and expectations. We can debate the extent to which these constraints are innate or acquired (and, consequently, absolute or tunable), but the unconstrained hypothesis is obviously a straw man.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

So too, I suspect, is **HECHT**'s externalization hypothesis. I believe he supplies us with this "thought exercise" primarily to demonstrate that, while formal descriptions derived from another domain (in this case, body dynamics) may initially seem to share constraints in common with our percepts, the mapping is ultimately limited and imprecise.

**HECHT**'s proposal of externalization, then, really points to the heart of the second question: *What* is internalized? HECHT delineates classes of regularities of the physical world that might be internalized. He then shows that most of these regularities are not consistently reflected in our perceptions (at least not when the regularities are "crisply" defined). But does this indicate a failure of the internalization hypothesis, or a failure to find a proper characterization of the internalized regularities?

With the possible exception of Gestalt principles, none of the candidate regularities were developed as descriptions of the psychological world. Newtonian physics are lovely abstractions, but they are seldom observable in our environment. Rather, objects are acted upon by forces that cannot be directly perceived (e.g., gravity, air resistance), and complex aspects of their kinematics (e.g., extended body motions) may not be penetrable by our visual system. Gibson (1979) reminded us that Euclidean geometry is not an appropriate description of our visual environment – rather, our world is filled with meaningful surfaces. So too, Newtonian dynamics do not describe the motions we perceive. If they did, a gyroscope would provide no wonder or amusement.

For better or worse, the human visual system is unlikely to conform to the simple, elegant solutions we seek to borrow from the domains of mathematics and physics. More likely, we will need to develop descriptions specific to perceptual phenomena, and to accept the challenging reality that our percepts reflect solutions derived from multiple, often competing, interpretive mechanisms. **SHEPARD** should be lauded for taking the bold step of offering an initial proposal. The other authors of this issue should be applauded for considering the limitations of Shepard's proposal, and proposing alternatives.

But the greatest work still lies ahead.

ACKNOWLEDGMENT

I thank Stevan Harnad and his staff for the opportunity to comment on this paper.

## The internalization of physical constraints from a developmental perspective

Horst Krist

*Department of Psychology, University of Zurich, CH-8032 Zurich, Switzerland.* **hkrist@genpsy.unizh.ch**

**Abstract:** Shepard's internalization concept is defended against Hecht's criticisms. By ignoring both Shepard's evolutionary perspective and the fact that internalization does not preclude modularization, Hecht advances inconclusive evidence. Developmental research supports Shepard's conclusion that kinematic geometry may be more deeply internalized than physical dynamics. This research also suggests that the internalization concept should be broadened to include representations acquired during ontogeny.

**[HECHT; SHEPARD]**

By discussing internalization without adopting a genetic or functional perspective, **HECHT** fails to do full justice to **SHEPARD**'s ideas. HECHT presents alleged evidence against Shepard's internalization hypothesis which appears largely beside the point. Would it have increased our ancestors' chances of survival if they had wired into their cognitive system a "spirit-level constraint" indicating that the surface of stationary liquids of low viscosity is horizontal with respect to the ground? And of what help could a prewired gravitational constant be, particularly if it only applies to falling objects and even then only if air resistance is negligible?

As far as gravitational acceleration is concerned, it is less important for an animal to know exactly how big a falling rock is, or how far away it is, than to know whether it is on a collision course and, if so, *when* it is time to run. A perceptual mechanism that computes "time to contact" (*tau*) from the dynamic information contained in the optical flow projected to the retina does indeed exist in various species (Lee & Thomson 1982; Wang & Frost 1992); and, at least on a rudimentary level, it seems to be present in human infants by their second month (Ball & Tronick 1971; Nánez 1988).

The functional relation between the rate of dilation of an image and the time to contact (*tau principle*) constitutes a physical regularity that holds for all situations in which an object is approaching an observer at a constant velocity. Despite focusing on innate components of an "intuitive physics," regrettably **HECHT** mentions the *tau principle* only in passing. Even more unfortunately, however, he completely ignores pertinent infancy research. **SHEPARD** himself (this volume) points to experimental evidence showing that infants exhibit sensitivity to geometrical constraints (e.g., continuity and solidity) much earlier than to dynamic constraints (e.g., gravity and inertia). Even as qualitative phenomena, gravity and inertia do not seem to be generally acknowledged by young infants (Kim & Spelke 1999; Spelke et al. 1994).

Thus, at least as far as the principles of gravity and inertia are concerned, it appears that one has to agree with **HECHT**'s (and **SHEPARD**'s!) negative conclusion concerning internalized constraints of physical dynamics. However, if one liberates the notion of internalization from the following two restrictions, things start to look different. First, it is not clear why one should reserve the internalization concept for constraints acquired phylogenetically. Second, the requirement that internalized constraints should be generally accessible or task-independent is much too restrictive.

Particularly in the domain of intuitive physics, a broader concept of internalization appears to be needed, one that includes the acquisition of internal representations during ontogeny (cf. Shepard 1984, pp. 431–32). Interpreting an acquired representation as internalized is meaningful if there is a functioning isomorphism (in the mathematical sense) between an aspect of the environment and a brain process that adapts the animal's behavior to it (Gallistel 1990, p. 3). Defining internalization in this way helps to integrate diverse branches of developmental and cognitive sciences. It further implies a functional perspective (Anderson 1996), one from which **HECHT**'s search for empirical evidence concerning the internalization of physical regularities appears misguided.

Internalization and modularization are by no means mutually exclusive, and this should be self-evident. Nevertheless, **HECHT** appears to take for granted that internalized principles can be revealed in perception, action, imagery, and problem-solving tasks alike whenever the situation is somehow underspecified. The premise of task-independence, however, does not generally hold, not even for those constraints that very young infants are sensitive to (von Hofsten et al. 1998; Spelke 1994). For example, Berthier et al. (2000) showed that 2-year-olds often act as if they know nothing about the impenetrability of solid objects (see also Hood 1995; Hood et al. 2000), although young infants have been shown to acknowledge the solidity constraint in various preferential-looking experiments (e.g., Baillargeon et al. 1985; Spelke et al. 1992). Conversely, infants exhibit sensitivity to inertia when reaching for moving objects (von Hofsten et al. 1998; Robin et al. 1996 ), but not in their looking preferences or search behavior if part of the trajectory is hidden from view (Spelke et al. 1994; 1995). Similarly, both children and adults have been shown to be sensitive to violations of inertia in their perceptual judgments (Kaiser et al. 1985a; 1992; Kim & Spelke 1999), but often fail to consider inertia in their explicit predictions and sometimes even in their actions (Krist 2001; McCloskey & Kohl 1983; McCloskey et al. 1983).

In light of these findings, it is conceivable that an inertia constraint is represented in certain perceptual-motor structures without being accessible for "higher" cognitive functions. Evidence for perceptual-motor structures representing physical laws also comes



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

from studies on the intuitive physics of projectile motion examined in the context of projecting balls from a horizontal board to hit targets on the ground (Krist et al. 1993; 1996). In these experiments, even young children exhibited detailed knowledge about the laws of motion in their throwing and kicking actions. For example, there was a virtually perfect correlation between the average throwing speeds produced by five- and six-year-olds and the speeds required for different height-distance combinations. That children's action knowledge is highly task-specific was not only revealed by the fact that they lacked explicit (judgmental) knowledge (Krist et al. 1993), but also by the finding that children up to the age of 9 or 10 ignored the release height when asked to operate a sling to project the balls (Krist 2000).

In the face of the pronounced task-specificity prevailing in the domain of intuitive physics, **HECHT**'s depiction of research findings concerning the water-level task, the C-tube problem, and the trajectory of thrown objects is not sufficiently elaborated upon. Regarding the issue of internalization, what can be usefully inferred from the fact that many people draw oblique water levels, wrong trajectories or misplaced velocity maxima in paper-and-pencil tests? Not much. Concerning the water-level task, McAfee and Proffitt (1991) found that people do not make any errors if they are led to evaluate the problem from an environment-related, instead of an object-related, perspective. Furthermore, Schwartz and Black (1999; Black & Schwartz 1996; Schwartz 1999) showed that children as well as adults are able to mentally simulate the dynamics of liquids in a glass being tilted for pouring. Their subjects exhibited implicit knowledge about the relation between glass width and the angle at which the liquid would spill when they tilted two glasses of different widths while pretending they were filled with water, but not when explicitly questioned which glass could be tilted further.

Concerning the C-shaped tube problem, **HECHT** himself notes that telling natural from anomalous trajectories is not a problem when subjects are shown visual animations. If I understand Hecht correctly, he considers erroneous predictions to be evidence against the internalization of inertia (Newtonian mechanics), and, at the same time, he considers the correct perceptual judgments to be evidence against the internalization of kinematic geometry. Both inferences are problematic, but the latter one particularly so. First, kinematic geometry does not prescribe any particular trajectory for objects exiting curved tubes, and there is no simplest path in this case. Second, even where such constraints or default solutions exist, our perception of object motions usually represents the actual trajectory rather than anything else. Our intuitive physics may be "contaminated" by constraints of kinematic geometry, but it certainly, and fortunately, also reflects our experience with the motions of real objects (cf. **SHEPARD**, this issue).

Among the arguments **HECHT** advances against the internalization hypothesis, those concerning his own externalization hypothesis are the weakest. His attempt to invalidate the former by trying to find evidence for the latter is futile. Far from being the opposite of internalization, externalization, as construed by Hecht, is equal to internalization (of body mechanics) plus erroneous "projection." Aside from this, it is disconcerting that Hecht does not scrutinize his own hypothesis as rigorously as he does **SHEPARD**'s.

In conclusion, **HECHT**'s target article is constructive in that it draws attention to a still undeveloped part of **SHEPARD**'s theory. Yet it is also destructive because it attacks the internalization concept without considering its place within the science of adaptive change.

ACKNOWLEDGMENT

I would like to thank Shem Barnett for his invaluable, professional help in editing an earlier draft of this commentary.

## The archeology of internalism

Martin Kurthen

*Department of Epileptology, University of Bonn, D-53105 Bonn, Germany.*
**martin.kurthen@meb.uni-bonn.de**
**http://www.meb.uni-bonn.de/epileptologie/**

**Abstract:** Behavioral regularities are open to both representationist (hence internalist) and non-representationist explanations. Shepard improvidently favors internalism, which is burdened with severe conceptual and empirical shortcomings. Hecht and Kubovy & Epstein half-heartedly criticize internalism by tracing it back to "unconscious" metaphors or by replacing it with weak externalism. Explanations of behavioral regularities are better relocated within a radical embodiment approach.
**[HECHT; KUBOVY & EPSTEIN; SHEPARD]**

***Internalism and representationism.*** "Internalism" can be said to be the doctrine wherein abstract features of (or "regularities in") the world are considered to be internal in current biological cognizers and where this internal presence is considered relevant to explaining the cognitive behavior of those beings. "Internalization" can be said to be the stronger claim (see **SCHWARTZ**) that those features of the world have been internalized *as a result of a specific process,* for example, the process of biological evolution. **SHEPARD** favors internalizationism and hence internalism, while **HECHT** and **KUBOVY & EPSTEIN (K&E)** question these doctrines. In what follows, I will assess only the weaker concept of internalism.

In a general sense, the issue of internalism merges with the issue of representationism, if the latter is taken as the claim that features of the world are *represented in* biological cognizers in a way that helps to explain the cognitive behavior of those beings. If a representation is the change that endures in an organism to mediate subsequent effects of an experience *after* that experience has ceased (Roitblat 1982, p. 354), then internality is a specified form of representationality in **SHEPARD**'s proof-bearing circadian-cycle example. But Shepard "internalizes" too easily and lightheartedly, and he seems to ignore the fundamental shortcomings of the issue of cognitive representationism (and hence internalism). **HECHT** and **K&E**, on the other hand, fail to radicalize their critique of internalism and thus factually play down the major problem of representation-based theories of cognition. In a nutshell, the problem is that internalism and representationism reverse the explanatory order: contrary to the representationists' assertions, representation cannot explain cognition, but cognition does (and has to) explain representation (Bickhard & Terveen 1995; Kurthen 1992).

***From literal internalism to as-if-internalism.*** **SHEPARD** has it that a behavior according to $x$ in the absence of $x$ implies the presence of specific "organizing principles" in the brain. That may be so, but it is not a proof of internalism, that is, not a proof that the relevant features of $x$ are literally represented in the organism. Take the example of the circadian rhythm. Some thousands of "clock neurons" in the suprachiasmatic nucleus of the hypothalamus form the principal circadian oscillator (Hastings 1997). This does not *entail* that the circadian rhythm is somehow literally represented in the brain; it just shows that there is a certain neurophysiological basis to circadian behavior. Even perfect scientific knowledge of (the neural correlate of) circadian behavior may at best justify the discourse of "as-if-internalism": the organism (the brain) proceeds *as if it had internalized* the external circadian rhythm. **SHEPARD** moves too easily from empirical data to the psychological entities postulated in theoretical explanation. Although it may seem that what we have been evolutionarily adapted to *is* what we have internalized, the issue of adaptation can (and should) be separated from the issue of internalism. What we have been adapted to is just what we are now able to handle and cope with (and adaptation may well be crucial for that ability), but that does not entail that we do so by means of literal internal representations of those features. The functioning of cerebral circadian oscillators, for example, is open to non-representationist interpretations. For example, Maturana and Varela (1987) have convincingly argued



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

that representation as a whole is an observer-dependent category which is not applicable to the level of cerebral processing proper. "Representation" is meaningful as part of a description of how certain cognitive systems succeed in acting and interacting on self-referential and social levels of communication. In this sense, representation is essential for social cognition, but it is not a property or mechanism that is literally efficient within brains (or minds).

So, although empirical data from cognition research are compatible with internalist interpretations, they do not entail internalism. But why ban this doctrine *tout court*? Because internalism, taken for granted as a basic premise of cognition theory, leads to a misconstruction of cognition as a whole (see Bogdan 1988; Kurthen 1992; Lakoff & Johnson 1999), taking representation as an *explanans* rather than *explanandum* within that theory. For even if features of the world were literally represented in brains/minds, these representations would not explain cognition. Representation as such establishes mere correspondence (or mapping relations), and this correspondence becomes a part of a cognitive system only if it is *processed according to its "mapping properties" in a way that makes this processing relevant to cognitive behavior.* In other words, representation presupposes cognition rather than explaining it (this problem of representationism is best illustrated in psychosemantics or the "symbol grounding problem" – see Bickhard & Terveen 1995; Kurthen 1992). In sum, literal representationism is nonexplanatory, and as-if representationism is merely illustrative with respect to basic individual cognition.

***Metaphors and premises.*** Hence, **K&E** are completely right in arguing (in other words) that the empirical data only support some sort of as-if internalism, that is, the statement that the organism behaves as if it has internalized some relevant physical property. But their rather crude psychological interpretation that the notion of internalization results from the unconscious application of misleading metaphors, in Lakoff and Johnson's (1999) sense, is too narrow-minded. It is not just seductive "unconscious" metaphors the scientist is fooled by; instead, it is well-identifiable theoretical premises and basic concepts of the representational approach in general that lead to the imprudent application of the internalization discourse. And a radical critique of internalism is more than just a therapeutic uncovering (and subsequent avoidance) of misleading pictures: it is a complete rejection or reversal of basic premises for reasons of explanatory power (see below on the archeology of internalism).

***Internalist externalism.*** **HECHT** is to be applauded for his attempt to move to an "opposite principle" in the face of some severe problems of internalism. But his "externalization" only superficially reverses the assumed cognitive procedure, thus leading to a sort of externalism within internalism. HECHT's externalism explains certain cognitive features as results of projections of properties of the organism into the world (instead of injections of worldly properties into the mind). But the fundamental issue of internalism is left untouched by this hypothesis, since in the externalist version, the mind (or brain, or representational system) will have to have some features internalized and represented, too: properties of the organism that are cognitively used in the mode of externalization still have to be represented or internalized in the system in order to become relevant parts of the cognitive process. In this approach, internalism is practically ineliminable because cognitive processes in general are still conceptualized in representationist terms. What is desirable is externalism without such internalization: that would be a radical, ecological, embodied interactionism, as described below.

***The (reverse) archeology of internalism.*** An adequate archeology of internalism will trace the notions of internality and representation back to the basic conceptual premises of the whole explanatory framework (not just its "metaphors," although the relation between the metaphors and the premises is another interesting field to study) – premises **SHEPARD**'s approach implicitly relies on and **HECHT** and **K&E**'s proposals are at least compatible with (that's why their critique of Shepard remains half-hearted). These premises have often been criticized (take again Lakoff & Johnson 1999 as the most recent and persuasive reference; the main philosophical source is Heidegger's 1927 work on *Sein und Zeit;* transl. Heidegger 1962; see Kurthen 1992). Briefly, three basic ideas can be identified:

(1) Confrontationism, as the view that the cognitive system and its environment are to be taken as two separate entities. In this view, the cognitive system is "confronted" with an external world it has to "build up" a relation to, and cognition is exactly the faculty that enables the system to bridge that gap;

(2) Representationism, as the idea that within the cognitive system, there are representations of aspects of the external world, that is, inner states that "stand for" an entity or feature of the world. Thus, representation is the very mechanism of cognition by which the gap opened by world-organism confrontation(ism) is bridged;

(3) Rationalism, as the belief that cognizing is best explained according to the model of *theorizing,* that is, the neutral attitude of directing one's look at an external object with theoretical interest, but without existential involvement. Not by pure chance, this is exactly the attitude of the (cognitive) scientist toward his/her scientific object.

Starting from these ideas, the well-known history of cognitive science runs not only to internalism, but to the whole of "orthodox" cognitive science known as "cognitivism," the "computational theory of cognition," the "representational theory of mind," and so on. But the story has continued: orthodoxy has had its crisis (due to its inability to model important aspects of cognition and its failure to solve the problem of cognitive representation, see above), from which connectionism, teleosemantics, situated cognition, dynamical systems theory, adaptive robotics, and so on, have arisen. Hence, the whole ecological and "embodiment" approach to cognition has arisen as an alternative framework, in which motivated embodied interaction is understood as the primary realm of cognition. In its radical guise, this approach literally reverses the above mentioned premises: it is held that the cognitive system and its environmental niche *together* form the unseparable, basic unit of cognition (thus rendering any re-presentation superfluous); that the cognitive system *acts* within this ecological unit rather than representing features of the world; and that cognizing is at root a motivated, ongoing, everyday practice rather than an aloof theorizing. In this approach, neurophysiological events (as the hypothalamic circadian oscillators) are not construed as internalizations or *representations* of features of the separated world, but rather as elements of an ecological cognitive system with a certain *role* for providing successful performance in a "circadian environment." In this account, representationism (and internalism) is a derivative, at best illustrative, discourse.

Thus, going back to **SHEPARD**, **HECHT**, and **K&E**, it is evident that internalism is:

(1) not warranted by empirical data, which are open to various – representationist and non-representationist – interpretations;

(2) not a fruitful concept since it is associated with all the unwelcome problems of orthodox cognitivism in the theory of cognition.

Hence, for a complete theory of cognition, internalism is the wrong credo. A reverse archeology of internalism will transfer the above mentioned backward-looking critique into a forward-looking research strategy in cognitive neuroscience, presently manifest mainly in the embodiment approach. In the long run, it is not promising to implicitly rely on the premise of internalism as part of cognitive "orthodoxy," as **SHEPARD** does; nor is it sufficient to criticize it half-heartedly as an approach seduced by unconscious metaphors (**K&E**), or to replace it by a superficial internalist externalism that keeps with confrontationism and representationism (**HECHT**). Radical ecological embodiment is the ideology (or metaphor) of choice.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

# Internalization of physical laws as revealed by the study of action instead of perception

Francesco Lacquaniti and Mirka Zago

*Human Physiology Section, Scientific Institute Santa Lucia, University of Rome "Tor Vergata," 00179 Rome, Italy.* **lacquaniti@caspur.it**
**m.zago@hsantalucia.it**

**Abstract:** We review studies on catching that reveal internalization of physics for action control. In catching free-falling balls, an internal model of gravity is used by the brain to time anticipatory muscle activation, modulation of reflex responses, and tuning of limb impedance. An internal model of the expected momentum of the ball at impact is used to scale the amplitude of anticipatory muscle activity.
**[BARLOW; HECHT; SHEPARD]**

Mental reflections of the outside world presumably are used not only to construct perceptual-cognitive representations, but also to guide our actions. However, little – if any – attention is paid to action-oriented representations in this series of articles. Our commentary is devoted to reviewing some evidence that indicates that physical laws may be internalized for our interaction with the environment in cases in which they are not overtly exploited for perception and cognition.

We first comment on the distinction drawn by **SHEPARD** between the internalization of kinematic geometry and that of physics. On the basis of studies of apparent motion perception, he suggests that geometry is more deeply internalized than physics. As far as interactions with outside objects are concerned, physics may not be easily dissociable from kinematic geometry, except under the special conditions of laboratory experiments involving virtual objects. Moreover, it may not be very useful for the brain to disassociate kinematics from dynamics. According to **BARLOW**, internalization means to copy a given environmental regularity internally; in other words, it implies that the brain uses an internal model that can mimic a specific property of the world. If the brain uses internal models that mimic interactions of our body with the external environment so as to prepare the appropriate motor commands, these models would be most useful if they encompassed dynamics in addition to kinematics. Actions require the production of muscle forces that are appropriately weighted for a given task; dynamic parameters of our own movements must be specified on the basis of the estimated dynamic parameters of interaction with external objects (such as mass, force, etc.). In the context of SHEPARD's metaphor that universal psychological laws should have a similar status as the invariant physical laws in general relativity, we suggest that brain representations of kinematics must include dynamics, just as in general relativity the geometry of space-time is affected by the presence of mass.

**HECHT** includes gravity among the prime candidates for the constraints that the nervous system might have internalized to disambiguate visual perception. However, he rejects the hypothesis of gravity internalization on the basis of his own studies of computer-simulated events of free-falling balls that indicate that observers are not very good at scaling absolute size and/or distance using acceleration cues. Aside from the issue that using virtual objects that lack the ecological impact of real objects may not tap into the same psychological domain, we would like to point out that even though gravity may not be internalized for constructing a visual representation of the object's motion, it does seem to be internalized for manual interception of the object. As HECHT himself states, the visual system has evolved not just to give us nice pictures of the world, but also to guide our actions.

Catching a free-falling ball is a good paradigm to study internalization of physics for action control. The role played by internal models of gravitational acceleration and object mass readily emerges (Lacquaniti & Maioli 1989a; 1989b). The planned interaction with the object requires that the physical parameters of the impact on the hand are accurately predicted. Time, location, and momentum of the impact must be estimated, and the activity of limb muscles accordingly controlled. Thus the hand must be placed so as to intercept ball trajectory, and limb rigidity must absorb ball momentum at the right time. Moreover, motor activity has to be anticipatory to overcome delays in the sensori-motor system. A crucial question is, how do we estimate the time remaining before contact (time-to-time contact, TTC)? According to a widely held hypothesis, visual signals alone provide the information to predict TTC (Lee 1980; Lee et al. 1983; McBeath et al. 1995; Rushton & Wann 1999; Savelsbergh et al. 1992). However, because acceleration is difficult to discriminate visually (Werkoven et al. 1992), it has been proposed that the brain may use an internal model of gravity to predict the acceleration of a falling object (Lacquaniti 1996; Lacquaniti & Maioli 1989a; Lacquaniti et al. 1993a; Tresilian 1993; 1999). As suggested by **SHEPARD** and **HECHT**, internalization may solve the problem of sensory under-specification, as resulting from ambiguous or incomplete sensory information.

This hypothesis has been tested in a series of studies involving catches of free-falling balls of identical external appearance but different mass, dropped from heights between 0.2 and 1.6 m (Lacquaniti & Maioli 1989a; 1989b; Lacquaniti et al. 1992; 1993a; 1993b). The electrical activity (EMG) of arm muscles was recorded. Figure 1 depicts the results for the EMG anticipatory activity of biceps. The time-to-onset of anticipatory activity relative to impact and the time course of the activity do not change with the height of fall, nor do they depend on the ball mass. Thus, the responses are precisely timed on impact time. A similar time-

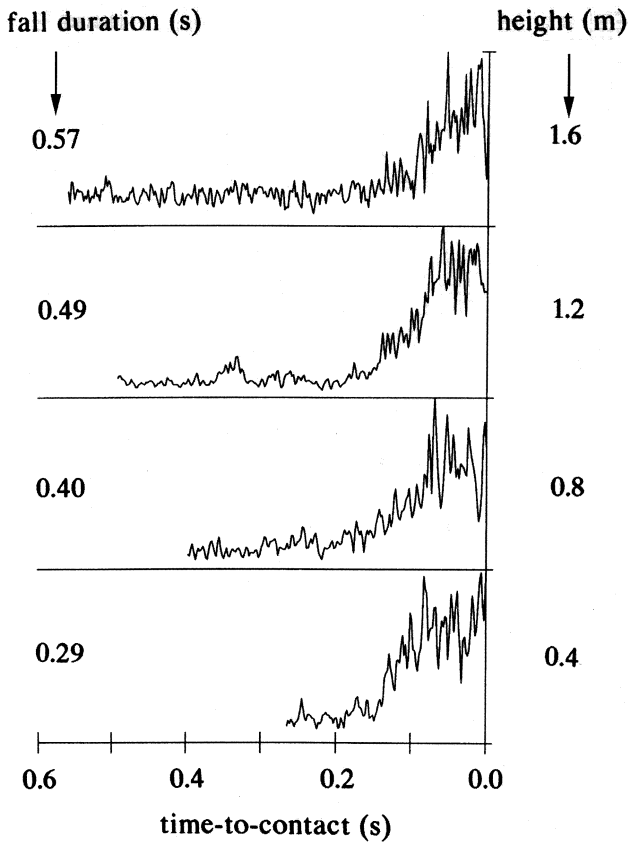



Figure 1 (Lacquaniti & Zago). Time course of EMG anticipatory responses for biceps muscle during catching. Traces correspond to the results obtained at the indicated heights of fall (right) and fall duration (left). EMG traces have been scaled to their maximum and aligned relative to impact time. Time axis indicates the time remaining prior to impact (TTC). The time to onset of anticipatory EMG build-up relative to the impact does not change systematically with height of fall (modified from Lacquaniti et al. 1993).


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

locking on impact is observed for the modulation of muscle reflex responses and for the changes in overall hand impedance (the mechanical resistance to an imposed displacement, see Fig. 2). Remarkably, motor preparation of reflex responses and limb impedance is correctly timed on impact even when blindfolded subjects are alerted of ball release by an auditory cue but have no real-time information about TTC (Lacquaniti & Maioli 1989b). The hypothesis that an internal model of gravity is used by the brain to time catching actions has recently been tested in micro-gravity as well (McIntyre et al. 1999). Astronauts caught a ball projected from the ceiling at different, randomized speeds both on ground (1g) and in-flight (0g). Motor activity started too early at 0g, with time shifts in accord with the internal model hypothesis. Apparently, they did not believe their eyes that told them the ball was traveling at constant velocity, but they behaved as if the ball was still accelerated by gravity.

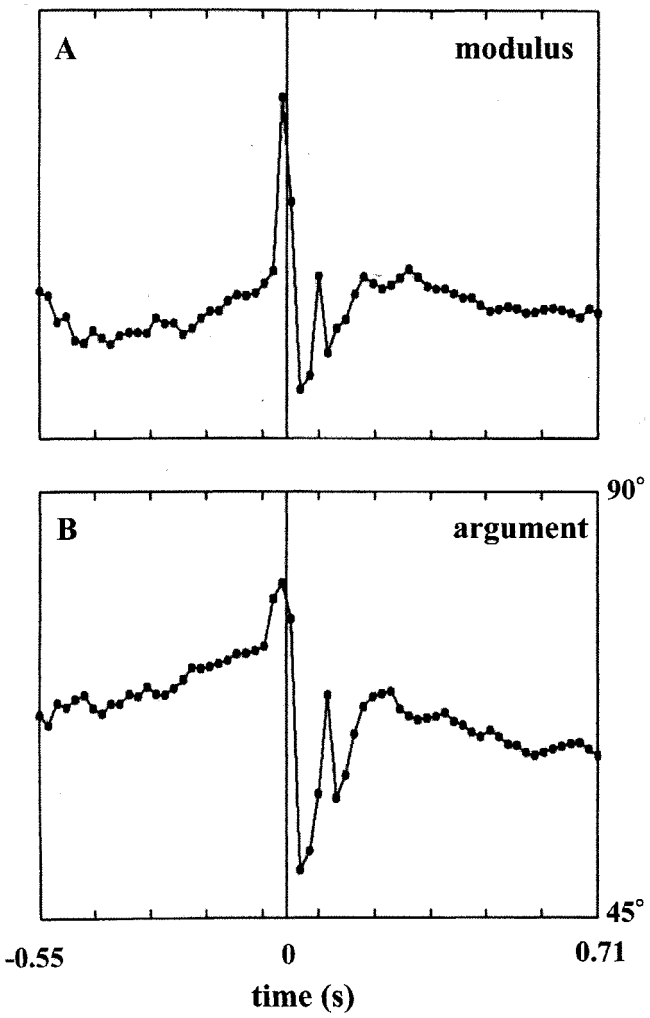



Figure 2 (Lacquaniti & Zago). Time course of the changes of end-point impedance during catching. Continuous, unpredictable perturbations were applied at the elbow joint by means of a torque motor, starting from ball release (time – 0.55 sec) through ball impact (time 0) and afterward. The time-varying values of stiffness and viscosity coefficients at the end-point were computed by cross-correlating input torque with output displacement. The modulus (arbitrary scale) and the argument of hand viscosity are plotted in A and B, respectively. A 0% argument corresponds to a horizontal vector pointing outward from the hand, whereas a 90% argument corresponds to a vertical, upward vector. Note that prior to ball impact, the magnitude of hand viscosity (and stiffness, not shown) increases significantly, while the direction of the viscosity vector rotates closer to the vertical, that is the direction of ball impact (modified from Lacquaniti et al. 1993b).

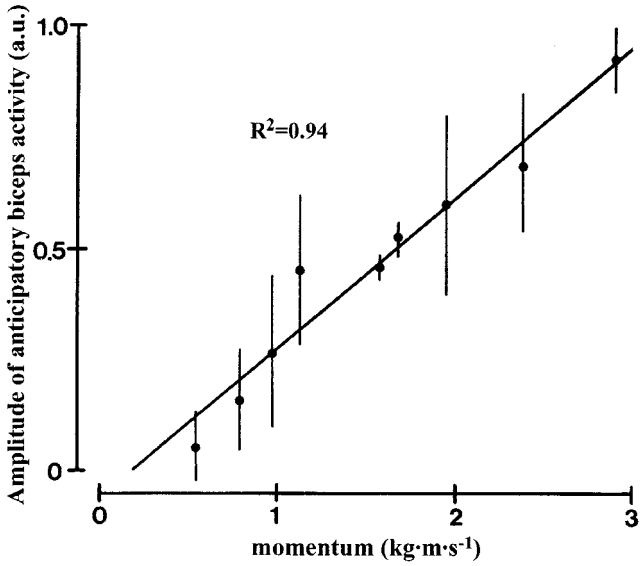



Figure 3 (Lacquaniti & Zago). Linear relation between the amplitude of biceps EMG anticipatory responses (mean value over the 50-msec interval preceding impact) and the momentum of the ball at impact time (modified from Lacquaniti & Maioli 1989a).

Catching studies also reveal that another dynamic parameter can be internalized, namely the predicted momentum at impact. Figure 3 shows that the amplitude of anticipatory muscle activity scales linearly with the expected momentum of the ball impact (Lacquaniti & Maioli 1989a). This was demonstrated using a factorial design, which involved the independent experimental manipulation of height of fall and mass of the ball. Thus, other kinematic or kinetic parameters could be excluded as putative control elements. In addition, it has been shown that, when the mass of the ball is unexpectedly changed, subjects scale their responses to the expected momentum.

In conclusion, we reviewed evidence that supports **SHEPARD**'s hypothesis that during our evolutionary development we have internalized environmental regularities and constraints. In particular, we showed that physical laws may be internalized for our interaction with the environment even in cases in which they are not overtly exploited for perception and cognition. Moreover, the internal models of dynamics we have considered for the task of ball interception also satisfy **BARLOW**'s criterion that the regularity must be turned to an advantage to have a biologically relevant value, as is well known to all fans of ball games.

## Extending Bayesian concept learning to deal with representational complexity and adaptation

Michael D. Lee

*Department of Psychology, University of Adelaide, SA 5008 Australia.*
**michael.lee@psychology.adelaide.edu.au**
**http://www.psychology.adelaide.edu.au/members/staff/michaellee/**

**Abstract:** While Tenenbaum and Griffiths impressively consolidate and extend Shepard's research in the areas of stimulus representation and generalization, there is a need for complexity measures to be developed to control the flexibility of their "hypothesis space" approach to representation. It may also be possible to extend their concept learning model to consider the fundamental issue of representational adaptation.
**[TENENBAUM & GRIFFITHS]**



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Two research areas in which Roger **SHEPARD** has made enormous contributions are stimulus representation (e.g., Shepard 1980) and stimulus generalization (e.g., Shepard 1987b). The Bayesian account of concept learning developed by **TENENBAUM & GRIFFITHS** (**T&G**) addresses both of these areas, providing a unifying consolidation of Shepard's representational ideas, and a natural extension of the "consequential region" approach to modeling generalization. A number of challenges and problems, however, remain for future research.

On the representational front, **T&G** demonstrate that their approach to building representations is sufficiently flexible to accommodate spatial, featural, and a range of other established approaches. While **T&G** note that part of the attraction of Shepard's (1987b) theory is that it assumes well-defined representational structures, their number game demonstrates the need for richer representational possibilities. By modeling stimulus representations in terms of prior distributions across an unconstrained hypothesis space, **T&G** develop an approach that may be sufficiently sub-conceptual (Smolensky 1988) to act as a useful unifying framework.

The price of (representational) freedom, however, is eternal (complexity) vigilance. The representational flexibility of the hypothesis space approach demands that the complexity of the representations be controlled. In the absence of some form of Occam's Razor, there is a danger that arbitrary stimulus representations can be constructed to solve particular problems, without achieving the substantive interpretability, explanatory insight, and generalizability that is the hallmark of good modeling.

What is required is a method for imposing priors on a hypothesis space that satisfy representational constraints in a parsimonious way. Following **T&G**, it seems plausible that representational constraints could be internalized through evolution, or learned on the basis of interaction with the world. Any source of information that offers adaptive advantage provides a candidate for representational refinement. The important point is that the representational priors must accommodate the constraints at an appropriate level of generality. Representations fail to serve their adaptive purpose if they do not generalize, and do not allow what has been learned (or internalized) in the past to be brought to bear on present concerns.

**T&G** are certainly aware of this challenge, as their discussion of the origin of representational priors indicates. Their general notion of developing a "vocabulary for a variety of templates" to tackle the challenge is an intriguing and promising one. One of the fundamental tools needed to pursue this undertaking, however, is a mechanism for assessing the complexity of arbitrary hypothesis space representations, and **T&G** are comparatively silent on this issue.

Fortunately, there are grounds for optimism. The Bayesian framework adopted by **T&G** is well suited to addressing issues of model complexity (Kass & Raftery 1995), and there have been recent attempts to develop Bayesian complexity measures for multidimensional scaling, additive clustering, and other approaches to stimulus representation subsumed under the hypothesis space approach (Lee 1999). The additive clustering analysis (Lee 2001) is particularly promising in this regard, since it gives measures that are sensitive to the "functional form" component of representational complexity (Myung & Pitt 1997), as will surely be required for the general hypothesis space approach. Indeed, given the formal correspondence between **T&G**'s Bayesian model and Tversky's (1977) ratio model, and the close relationship of the ratio model to the contrast model that underpins additive clustering, some of the groundwork has already been laid.

In terms of stimulus generalization and concept learning, the model developed by **T&G** constitutes an impressive extension of Shepard's (1987b) approach, particularly through the introduction of the size principle. Their Bayesian formulation seems to capture important capabilities of human learning that are not obviously present in discriminative learning models such as ALCOVE (Kruschke 1992).

One issue that **T&G** do not substantially address is representational adaptation resulting from learning. A fundamental problem for any adaptive system with a memory is: how should established representations be modified on the basis of experience? The Bayesian account of concept learning involves the interaction of data-driven (perceptive) and knowledge-driven (apperceptive) components, and so is well placed to deal with this issue. Studies of learned categorical perception that measure the effects of concept learning on human mental representations (Goldstone et al. 2001) could provide one source of empirical data to guide theoretical development.

Ultimately, addressing the issue of adaptation requires an understanding of the way in which perceptive and apperceptive processes interact across different learning episodes and time scales. The Bayesian concept learning model modifies its representations to learn a particular concept from a small number of stimuli, but the permanence of these modifications is not clear. If a new concept is subsequently learned across the same stimulus domain, what is the effect of previous learning? Do the priors on the hypothesis space revert to their original state, or do they assume a different distribution that is partly influenced by the learned concept? In some cases, it seems likely that the representations will be unchanged. It would come as no surprise if human performance on repeated versions of the number game were shown to be independent of each other. For particularly salient concepts, or for conceptual relationships that are continually reinforced over time, however, there is a much stronger argument for change. On evolutionary time scales, the argument that representations have adapted in response to ancestral experience is compelling. Extending the Bayesian model of concept learning to balance the use of representations in learning with the use of learning in representation-building should be a focus of future research.

Finally, it may be worth some effort exploring the relationship between the Bayesian approach of **T&G**, and the "fast and frugal" approach to cognitive modeling (Gigerenzer & Todd 1999b). In discussing a related Bayesian model of prediction (Griffiths & Tenenbaum 2000), the same authors have argued that humans do not actually perform the Bayesian calculations specified by their model, but instead apply a simple heuristic that approximates the outcomes of these calculations. It would be interesting to know whether **T&G** hold the same view in relation to their model of concept learning and, if so, what sorts of heuristics they believe are likely to be involved.

ACKNOWLEDGMENTS

I wish to thank Daniel Navarro and Chris Woodruff for helpful comments.

## Representation of basic kinds: Not a case of evolutionary internalization of *universal* regularities

Dennis Lomas

*Department of Theory and Policy Studies, Ontario Institute for Studies in Education, Toronto, Ontario, M5S 1V6, Canada.* **dlomas@oise.utoronto.ca**

**Abstract:** Shepard claims that "evolutionary internalization of universal regularities in the world" takes place. His position is interesting and seems plausible with regard to "default" motion detection and aspects of colour constancy which he addresses. However, his claim is not convincing with regard to object recognition.
**[SHEPARD]**

**SHEPARD** makes a convincing case for "evolutionary internalization of universal regularities in the world" with regard to "default" motion detection and to the aspects of colour constancy he addresses. His (provisional) attempt to apply the same principles to object recognition is not convincing. (I address the first five paragraphs of the sect. 1.10, "Formal characterization of generalization based on possible kinds.")



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

The cognitive categorization of objects which **SHEPARD** describes and addresses is quite sophisticated. (See the third paragraph of sect. 1.10.) It involves precise categorizing of objects according to, as he says, basic kinds. A dog, but not a statute of a dog, is recognized as a dog. (This is my example.) Such precise categorization is beyond the reach of perceptual capacities. Stimulus similarity of objects from disparate basic kinds causes faulty recognition because perceptual systems are tied to stimulus configurations. The evolutionary success of mimicry and other deceptions is ample testimony to the intrinsic limitations of perception to pick out basic kinds accurately.

**SHEPARD** may hold that the cognitive resources of an individual involved in the type of categorization which he has in mind are not restricted to perceptual resources. At least he seems to be going in this direction. (See the first paragraph of the section.) Involvement of a broad range of cognitive resources increases the likelihood of precise recognition according to basic kinds. We have become better at differentiating the real from the fake. Collection of evidence and logic have aided us along this path. (For example, I can accurately infer that I am looking at a statue of a dog, not a real dog, from the fact that the thing has not moved a hair's breadth in five minutes.) However, many cognitive capacities which are involved in object recognition do not derive from internalization of universal regularities. Belief systems, for example, are enormously plastic. At one time, many people believed they saw Zeus when they looked at a cloud containing the shape of a bearded head. That belief does not occur very much anymore. Fluidity of beliefs seems to be a prerequisite for steady progress toward precise identification (even regarding basic kinds). In contrast, perceptual mechanisms, which are more likely to involve evolutionary internalization of universal regularities, are unreliable.

Another concern arising from **SHEPARD**'s proposal is this: not all of those things which Shepard calls "basic kinds" are universal. Animal species arise, decline, and disappear, others arise, and so on. Therefore, the cognitive mechanisms which induce recognition of a specific animal (e.g., a lion) have not in general become tuned to universal regularities, but to contingent regularities. This consideration, it would seem, blocks application of Shepard's theory to representation of basic kinds. I have touched on one way only in which contingency blocks Shepard's attempt to theoretically capture representation of basic kinds. In order to recognize objects in this world adequately, the tie of cognitive systems to universal regularities must be strictly limited. This applies to perceptual mechanisms as well, insofar as they induce object recognition, because, generally speaking, the objects at issue are not universal.

In making his case, **SHEPARD** talks in terms of connected regions in representational space which correspond to basic kinds. (See second paragraph of the section.) These regions are constructed by an individual's judgement of similarity of consequences. This way of describing things does not seem to diminish my criticisms. Many types of consequences in ordinary environments are just as contingent and fluid as many basic kinds, requiring that the mechanisms which underlie judgements of similarity of consequences must be substantially cut loose from universal regularities. As an example of the contingency of consequences, consider the contrast between the consequence of encountering a live rattlesnake and that of encountering a dead one. In order to differentiate between these two consequences, cognition cannot be completely tied even to semi-permanent regularity, such as the size and colouration of the rattle snake.

If my criticism holds, **SHEPARD**'s mathematical project (described in his conclusion) is in jeopardy with respect to representations of kinds of objects. Because these representations in the main are not tied to universal regularities, mathematical models which link these representations to universal regularities are bound, in general, to have only limited scope. Of course, sophisticated mathematics can be, and is, fruitfully used to model representations of basic kinds, but not in general to tie these representations to universal regularities; instead mathematics can be, and is, often used to model representations which are tied to contingent, even radically contingent, regularities.

## Three deadly sins of category learning modelers

Bradley C. Love

*Department of Psychology, The University of Texas at Austin, Austin, TX 78712.* **love@psy.utexas.edu** **http://www.love.psy.utexas.edu/**

**Abstract:** Tenenbaum and Griffiths's article continues three disturbing trends that typify category learning modeling: (1) modelers tend to focus on a single induction task; (2) the drive to create models that are formally elegant has resulted in a gross simplification of the phenomena of interest; (3) related research is generally ignored when doing so is expedient. **[TENENBAUM & GRIFFITHS]**

***Overview.*** **TENENBAUM & GRIFFITHS**'s (henceforth **T&G**) article continues three disturbing trends that typify category learning modeling: (1) modelers tend to focus on a single induction task, which drastically limits the scope of their findings; (2) the drive to create models that are formally elegant has resulted in a gross simplification of the phenomena of interest and has impeded progress in understanding how information is represented and processed during learning; (3) related research on the role of theories, prior knowledge, comparison, analogy, similarity, neurospychology, and cognitive neuroscience is generally ignored when doing so is expedient. These three shortcomings are all interrelated and mutually reinforcing.

***Induction tasks: The unwarranted assumption of universality.*** **T&G** exclusively focus on how subjects generalize from positive examples of a single target concept. This learning mode can be characterized as unsupervised learning under intentional conditions because subjects are aware that they are in a learning task and all of the training examples are from the same target concept (i.e., discriminative feedback or supervision is not provided). **T&G** ignore other learning modes such as classification learning, inference-based learning, and unsupervised learning under incidental conditions. This oversight is important because the ease of acquiring target concepts differs greatly depending on which of these learning modes is engaged. For example, inference-based learning is more efficient than classification learning when the task is to acquire two contrasting categories that are linearly separable (i.e., there is a linear decision boundary in representational space that separates examples of categories "A" and "B"), but is less efficient than classification learning for nonlinear category structures (Love et al. 2000; Yamauchi & Markman 1998). Recent work in my lab (in preparation) demonstrates strong interactions among all four of the learning modes mentioned above.

Given these interactions between learning problems and learning modes, focusing exclusively on a single learning mode is problematic to any theory that intends to explain category learning and generalization in any comprehensive sense. Currently, the category learning literature focuses on classification learning, which limits the field's ability to construct general theories of category learning. This narrow focus also raises concerns of ecological validity because, as Yamauchi and Markman (1998) have demonstrated, classification learning does not support inference (i.e., predicting an unknown property of an object from a known category). Ostensibly, inference is a major use of categories. The current fascination with classification learning can be traced back to Shepard et al.'s (1961) seminal studies which, oddly enough, are not considered by **T&G**.

***It doesn't have to be pretty to be beautiful.*** **T&G** invoke evolutionary arguments, but higher-level cognition is probably best regarded as a "hack" involving multiple learning, memory, and control systems – many of which were probably co-opted or developed rather recently in our evolutionary history. The growing



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

consensus in the memory literature is that memory is not unitary, but instead involves multiple systems (e.g., semantic, episodic, declarative, etc.) that operate in concert (Cohen & Eichenbaum 1993; Squire 1992). Some category learning researchers have recently embraced this idea with multiple system learning models (Ashby et al. 1998; Erickson & Kruschke 1998). Even work that argues against the multiple systems approach (e.g., Jacoby 1983; Roediger et al. 1989) emphasizes the importance of how a stimulus is processed at encoding. In light of these results, the search for a universal (monolithic) theory of learning seems at best misguided.

In general, the field has been attracted to models that are rather abstract and that can be construed as optimal in some sense (e.g., Ashby & Maddox 1992; Nosofsky 1986). Unfortunately, it seems unlikely that an ideal observer model (of the type commonly deployed in psychophysics research) can be applied to understanding human category learning in any but the most trivial sense (e.g., to understanding Boolean concept acquisition via classification learning as in Feldman 2000). Clearly, theories cannot be formulated at an abstract informational level because learning modes that are informationally equivalent (e.g., inference-based vs. classification learning; intentional vs. incidental unsupervised learning) lead to different patterns of acquisition.

What is needed are models that account for the basic information processing steps that occur when a stimulus is encountered. Current category learning models err on the side of the abstract (neglecting processing) and do not make allowances for basic processing constraints (e.g., working memory limitations). Accounting for basic processing mechanisms will lead to insights into the nature of category learning. For example, SUSTAIN's (SUSTAIN is a clustering model of category learning; see http://love.psy. utexas.edu/ for papers) successes are largely attributable to its characterization of how and when people combine information about stimuli.

**T&G** move even farther away from issues of processing and representation. Contrary to appearance, their framework lacks explanatory power. In their model, many layers of representation and processing (e.g., constructing hypothesis, resolving conflicting hypotheses, updating model memory) are collapsed into a hand-coded hypothesis space. This framework makes it impossible to address important issues like whether people are interpolating among exemplars, storing abstractions, applying rules, constructing causal explanatory mechanisms, and so forth, because all possibilities are present and lumped together. Additionally, there is little psychological evidence that humans perform Bayesian inference. Instead, humans tend to focus on the most likely alternative, as opposed to performing a weighted (by probability) summation over all alternatives and the corresponding values (Murphy & Ross 1994).

***Let's learn from others.*** Category learning modelers show an alarming disregard for research in related literatures. I will leave it to the other commentators to castigate **T&G** for dismissing the last twenty years of research in analogy and similarity based on what amounts to a thought experiment. It suffices to say that relations are not features and that features and relations are psychologically distinct (Gentner 1983; Goldstone et al. 1991).

While many other category learning modelers are guilty of not making contact with related work (e.g., the role of prior knowledge in learning), **T&G** actually fail to make contact with other models of category learning by example. T&G dismiss other models of category learning in their "Alternative approaches" section 3.3, without addressing any of the data supporting these "alternative" models.

## Tribute to an ideal exemplar of scientist and person

Dominic W. Massaro

*Department of Psychology, Social Sciences II, University of California-Santa Cruz, Santa Cruz, CA 95064.* **massaro@fuzzy.ucsc.edu**
**http://www.mambo.ucsc.edu/psl/dwm/**

**Abstract:** Roger Shepard's creativity and scientific contributions have left an indelible mark on Psychology and Cognitive Science. In this tribute, I acknowledge and show how his approach to universal laws helped Oden and me shape and develop our universal law of pattern recognition, as formulated in the Fuzzy Logical Model of Perception (FLMP).
**[SHEPARD; TENENBAUM & GRIFFITHS]**

It is fitting that *BBS* should sponsor a forum on Roger **SHEPARD**'s seminal contributions to the understanding of mind and behavior. His work has always been earmarked by creativity, innovation, and relevance. All of this by a most unassuming person. I shared a plane ride with him after he had just been awarded the Presidential Medal of Science at the Whitehouse. He was as interested, curious, and supportive as always, without exposing any hint of the great honor he had just received.

Laws are lofty targets out of reach by most of us. **SHEPARD** created a law imposing order on one of the oldest problems in experimental psychology. How do we account for behavioral responses to stimuli that are similar but not identical to a stimulus that has been previously shown to be informative? Generalization was not simply a matter of failure of discrimination (Guttman & Kalish 1956); and what function could possibly describe the myriad conglomerate of findings across organisms, stimuli, tasks, and so on?

**SHEPARD**'s solution was to enforce a distinction between the physically measured differences between stimuli and the psychological differences between those same stimuli. In many respects, this move was simply an instantiation of his general dissatisfaction with the prevalent behaviorism of the era. SHEPARD imposed order on unorderly data by making this distinction. His analysis of a broad range of data across different domains produces a highly consistent and universal function that describes generalization. When generalization between stimuli is predicted from distances between points in a psychological space, the resulting generalization function is exponential.

We have proposed the fuzzy logical model of perception (FLMP; Oden & Massaro 1978) as a universal law of pattern recognition (Massaro 1996; 1998). The assumptions central to the model are: (1) persons are influenced by multiple sources of top-down and bottom-up information; (2) each source of information is evaluated to determine the degree to which that source specifies various alternatives; (3) the sources of information are evaluated independently of one another; (4) the sources are integrated to provide an overall degree of support for each alternative; and (5) perceptual identification and interpretation follows the relative degree of support among the alternatives. In a two-alternative task with /ba/ and /da/ alternatives, for example, the degree of auditory support for /da/ can be represented by $a_i$, and the support for /ba/ by $(1-a_i)$. Similarly, the degree of visual support for /da/ can be represented by $v_j$, and the support for /ba/ by $(1-v_j)$. The probability of a response to the unimodal stimulus is simply equal to the feature value. For bimodal trials, the predicted probability of a response, P(/da/) is equal to:

$$P(|da|) = \frac{a_i v_j}{a_i v_j + (1-a_i)(1-v_j)} \qquad (1)$$

In the course of our research, we have found that the FLMP accurately describes human pattern recognition. We have learned that people use many sources of information in perceiving and understanding speech, emotion, and other aspects of the environment. The experimental paradigm that we have developed also al-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

lows us to determine which of the many potentially functional cues are actually used by human observers (Massaro 1998, Chapter 1). This paradigm has already proven to be effective in the study of audible, visible, and bimodal speech perception.

**SHEPARD**'s innovative analysis of the process of generalization anticipated exactly the strategy that we have taken in our theoretical development. In the application of our universal principle of pattern recognition, we necessarily make a distinction between information and information processing. Information in our view corresponds to how informative some source of information is, in terms of the degree of support that it provides different alternative interpretations. Our universal law of information processing concerning how multiple sources of ambiguous information are integrated, is revealed only when the information available to the perceiver is taken into account. This is exactly analogous to SHEPARD's distinction between the physical and the psychological properties of stimuli. Without first measuring the information, the information processing is not apparent. Our law of pattern recognition is most clearly seen when the information from each source is explicitly defined. Instantiated in this manner, the graph of the theory and data resulting from the influence of two sources of information follows the shape of an American football. This outcome reflects the fact that both sources of information are influential and are combined in an optimal manner. Shepard's groundwork made it easier to understand how people impose meaning on a world with multiple sources of ambiguous information.

For a critic of categorical perception, it is encouraging to analyze this phenomenon in terms of **SHEPARD**'s law of generalization. As formalized in his law, generalization is *not* a failure of discrimination. What researchers usually interpret as categorical perception is really nothing more than generalization. What has to be emphasized repeatedly is that categorical perception is not a failure to discriminate. As anticipated by Shepard, we simply generalize from one experience to similar experiences and treat them in similar ways. If we have a speech continuum between /ba/ and /da/, it should not be surprising that we tend to treat instances within a category as more similar to one another than instances between categories in both categorization and discrimination tasks.

**SHEPARD**'s law of generalization also offers a potential clarification of our prototype representations that mediate pattern recognition. We have defined these representations as summary descriptions of the ideal feature value for each feature of each test alternative. Given that the ideal values seldom occur in the stimulus to be recognized, how is the goodness-of-match determined? Normally, we would have to claim that some additional representation of the distribution of feature values is included in the summary description. With Shepard's principle, however, we could assume that only the ideal value is maintained in memory; the truth value indicating goodness-of-match decays exponentially with the psychological distance between the feature in the stimulus and its ideal value in memory.

**SHEPARD**'s approach and the article by **TENENBAUM & GRIFFITHS** highlight the theoretical importance of optimality of efficiency of information processing. We have shown that the FLMP is an optimal algorithm for combining multiple sources of information. Thus, the FLMP can be used to assess integration efficiency. As can be seen in Equation 1, the auditory and visual sources of support are multiplied to give an overall degree of support for each response alternative. The value $a_i$ representing the degree of auditory support is assumed to be the same on both unimodal auditory and bimodal trials. The same is true for the visual support. This property and the multiplicative integration rule, followed by the relative goodness rule (RGR), entail the process to be optimal and thus maximally efficient (see Massaro 1998, pp. 115–17; Massaro & Stork 1998). Empirical results in a variety of domains can therefore be assessed to determine if utilization of the two sources of information was optimal or maximally efficient (see Massaro & Cohen 2000).

Shepard (1957), along with Clark (1957), Luce (1959), Selfridge (1959), and Anderson (1981), envisioned the importance of relative rather than absolute goodness-of-match in determining selection of an alternative. This is particularly critical to pattern recognition in situations with a varying number of sources of information. Our research has consistently demonstrated that two sources of information lead to more reliable categorization than either one alone. In the context of the FLMP, integration involves a multiplicative combination of their two respective truth values (Oden 1977). Because the truth values are less than one, their multiplicative combination will necessarily be less than the value of either one individually. As an example, consider the case in which audible and visible speech each support /ba/ to degree .7 and /da/ to degree .3. On bimodal trials, the total degree of support for /ba/ would be $.7 \times .7 = .49$. The degree of support for /da/ would be $.3 \times .3 = .09$. On unimodal trials, the degree of support would be .7 for /ba/ and .3 for /da/. If the decision is based on absolute support, then the likelihood of a /ba/ judgment would necessarily be greater on unimodal (.7) than bimodal (.49) trials – an incorrect prediction. This observation has been taken as an inconsistency in fuzzy logic by Osherson and Smith (1981; 1982), but neither their analysis nor Zadeh's reply (1982) considered the role of using the relative degree of support for the different alternatives. However, the RGR normalized the predicted outcomes in the FLMP, and accounts for the data. If relative goodness is used, as in our example, then the total degree of support for /ba/ on bimodal trials would be .49 divided by $(.49 + .09) = .85$ – a correct prediction.

## What's within? Can the internal structure of perception be derived from regularities of the external world?

Rainer Mausfeld

*Institute for Psychology, University of Kiel, D-24098 Kiel, Germany.*
**mausfeld@psychologie.uni-kiel.de**

**Abstract:** Shepard's approach is regarded as an attempt to rescue, within an evolutionary perspective, an empiricist theory of mind. Contrary to this, I argue that the structure of perceptual representations is essentially codetermined by internal aspects and cannot be understood if we confine our attention to the physical side of perception, however appropriately we have chosen our vocabulary for describing the external world. Furthermore, I argue that Kubovy and Epstein's "more modest interpretation" of Shepard's ideas on motion perception is based on unjustified assumptions. **[KUBOVY & EPSTEIN; SHEPARD]**

Nativist-empiricist theories of mind could be conceived of as being based on the conception that the mind is endowed with a rich and innately specified internal structure, which, however, is determined entirely by experience, albeit experience as generalised within evolutionary history. On this account, **SHEPARD** is a nativistic empiricist. In his emphasis on phylogenetic experience he follows Spencer who, in his *Principles of Psychology* (1881), postulated a "continuous adjustment of internal relations to external relations." According to Spencer, the structure of the mind is the "result from experiences continued for numberless generations," whereby the "uniform and frequent of these experiences have been successively bequeathed" in the process of evolution. James (1890) lauded this as a "brilliant and seductive statement" that "doubtless includes a good deal of truth." It founders, however, according to James, "when the details are scrutinised, many of them will be seen to be inexplicable in this simple way."

**SHEPARD**, in contrast to Spencer, has made very specific proposals about the kind of external regularities that, in his account, have molded the structure of internal representations. He clearly recognises the explanatory vacuum caused by the disregard for postulating, within explanatory frameworks, specific internal structures adequate to the task of explanation. (Such disregard, which is characteristic of empiricist theories of mind, still prevails, in various modern disguises, in much of current thinking about percep-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tion.) SHEPARD rightly acknowledges that we have to assume a rich internal structure of the perceptual system in order to account for the relevant facts. He thus draws our attention to a core problem of perception theory, viz., to understand the structural form of internal representations. To this end, SHEPARD extends the approach of ecological physics to further kinds of abstract mathematical descriptions of external regularities, which he then uses as heuristics for exploring the structure of internal representations. His grand perspective on the *Evolution of a mesh between principles of the mind and regularities of the world* (1987a) doubtless includes a good deal of truth, notwithstanding the problems that his notions of "regularity" and "internalisation" are faced with when one attempts to understand them beyond their meaning in ordinary discourse. SHEPARD's more extensive (non-Darwinian) claim that there is an "evolutionary trend toward increasing internalization" (1987a, p. 258) and that by internalising more and more physico-geometrical regularities the fitness of a species is increased, is not easy to assess and would hardly be maintained in other areas of biology. Fortunately, issues of evolutionary internalisation do not bear any immediate relevance with respect to perceptual theory, because here, as elsewhere in biology, a satisfactory *ahistorical* account for a functional structure does not ipso facto suffer from some kind of explanatory deficit (cf. Fodor 2000). It seems to me that the role that the concept of internalisation plays in SHEPARD's account resembles the role that mechanisms of association play in standard empiricist approaches to the mind, viz., it acts as a kind of general multi-purpose acquisition device for building up mental structure.

What appears to me to be more problematic than the metatheoretical discourse about internalisation is **SHEPARD**'s extreme physicalistic stance. In SHEPARD's view the structure of internal representations is determined predominantly by regularities of the external world, whereas no essential explanatory importance is attached to those aspects of the internal conceptual structure of perception that do not mirror external regularities, or to internal constraints of the cognitive architecture. Shepard (1984, p. 431; 1987a, p. 269) seems to think that constraints on the principles of the mind that do not have an external origin are merely arbitrary. Naturally, they must appear arbitrary if one slices the nature of perception according to external physical regularities, thus succumbing to the *physicalistic trap* in perception theory (cf. Mausfeld, in press).

Evidently, there is sufficient overlap between regularities of the world and the structure of internal representations. However, from this global property, which pertains to the entire organism, it does not follow that the representational structure of specific subsystems is predominantly determined by specific features of the environment. With respect to internal codes, equivalent classes of sensory inputs are held together by the conceptual structure of our perceptual system, rather than by the structure of the physical environment itself. The given conceptual structure that is part of our biological endowment is based on concepts that are not expressible as "natural kinds" or abstract regularities of the external physical world. This is evident for internal perceptual concepts such as "edible things and nutrients" or "emotional states of others." In other cases, such as the *internal* concept "surface colour," it may be less obvious that it defies definition in terms of a corresponding physical concept (even in the sense of the latter providing necessary and sufficient conditions for the former). Rather, it has its own peculiar and yet-to-be identified relation to the sensory input and depends intrinsically, in an idiosyncratic way that cannot be derived from considerations of external regularities, on other *internal* codes, say, for perceived depth or figural organisation. (All the same one might be able to concoct some Panglossian post hoc story in terms of external regularities for each specific case, but nothing about an external origin would be implied by this.) The structure of internal representations, as Gestalt psychology and ethology have already provided ample evidence for, is shaped not only by regularities of the external world. Rather, internal representations have to fit into the entire conceptual structure of the perceptual system, including its two fundamental interfaces, viz., the interface with the motor system and that with the higher cognitive system, where meanings are assigned in terms of "external world" properties.

**SHEPARD** has reinvigorated psychological inquiries into the structural form of mental representations. Such inquiries inevitably lead back to the core problem of perception theory, viz., to understand the internal conceptual structure of perception. This problem, however, cannot be solved or dodged by exclusively referring to physico-geometrical or statistical regularities of the external world and by assuming that the rich structure is imprinted on the mind of the perceiver almost entirely from without.

While **SHEPARD** seems to accept internal structure only to the extent that an external origin dignifies it with a stamp of approval, as it were, **KUBOVY & EPSTEIN** relapse altogether into a wariness about postulating specific internal structures. They refer to a distinction, widespread in empiristic approaches to the study of the mind, between what they call a "measurement model of perception" and assumptions of "invisible internal principles." Because they do not want to lodge the principles that are part of a successful explanatory account in the mind of the percipient, they propose what they call a "more modest interpretation," according to which we can, instead of talking about internal principles, only say that the visual system proceeds *as if* it obeys internal principles. Thus, they implicitly make the distinction between evidence for an explanatorily successful theory and evidence for the "psychological reality" of the principles to which this theory refers. Even if SHEPARD's investigations on motion perception provided, at the level of description on which he is working, a successful explanatory account – both in range and depth – of an important class of facts, it would still lack, in KUBOVY & EPSTEIN's view, "psychological reality." This is a highly questionable and unjustified distinction, which would hardly be of interest elsewhere in the natural sciences. A similar request for a "more modest interpretation" in physiology with respect to, say, the idea that "pattern cells in area MT employ the assumption of smoothness in their computations of motion" (Hildreth & Koch 1987, p. 508) would justly be regarded as being without any theoretical interest. In perception theory, as in other fields of the natural sciences, we proceed by attributing to the system under scrutiny whatever serves our explanatory needs. Ascribing inner structure to the perceptual system is not some mysterious ontological commitment, but a case of an inference to the best explanation (subject to further inquiry and open to change). There are (aside from metaphysical issues) no ontological questions involved beyond what is stated by the best current explanatory account. The distinction that KUBOVY & EPSTEIN make is an instance of what Chomsky (2000) has identified as an odd dualism of explanatory principles between psychology and the rest of the natural sciences. Such a dualism, which is expressed in KUBOVY & EPSTEIN's emphasis on "measurement theories of perception," will impede asking, as SHEPARD does, fruitful questions about the "invisible internal principles," – a natural concern, it seems, for inquiries into the nature of the perceptual system.

## Probabilistic functionalism: A unifying paradigm for the cognitive sciences

Javier R. Movellan and Jonathan D. Nelson

*Department of Cognitive Science, University of California San Diego, La Jolla, CA 92093.* **movellan@mplab.ucsd.edu**
**http://www.mplab.ucsd.edu/** **jnelson@cogsci.ucsd.edu**
**http://www.cocsci.uscs.edu/~jnelson/**

**Abstract:** The probabilistic analysis of functional questions is maturing into a rigorous and coherent research paradigm that may unify the cognitive sciences, from the study of single neurons in the brain to the study of high level cognitive processes and distributed cognition. Endless debates about undecidable structural issues (modularity vs. interactivity, serial vs. parallel processing, iconic vs. propositional representations, symbolic vs.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

connectionist models) may be put aside in favor of a rigorous understanding of the problems solved by organisms in their natural environments. [SHEPARD; TENENBAUM & GRIFFITHS]

**TENENBAUM & GRIFFITHS'** (T&G's) paper on **SHEPARD**'s law of generalization is a beautiful example of the most exciting and revolutionary paradigm to hit the cognitive sciences since connectionism. We call this paradigm "probabilistic functionalism" for its focus on functional rather than structural questions and for its reliance on the machinery of probability and information theory. Probabilistic functionalism traces back to Brunswik (1952) and finds modern articulators in Marr (1982), Anderson (1990), and Oaksford and Chater (1998).

Suppose an organism was rewarded for pecking in response to a red key. Would the organism generalize the pecking behavior in response to a purple key? Shepard (1987b) observed that in a very wide variety of experiments, the degree of generalization to new stimuli is an exponential function of the perceived similarity between the old and new stimuli (see Fig.1, top). Under the dominant structural paradigm, one would typically approach this result by formulating mechanistic models of information processing, for example, connectionist networks, that exhibit the law. Yet even under the unrealistic assumption that one can uniquely specify the mechanisms of the mind, the structural approach ultimately fails to answer a critical question: why does the mind use such mechanisms?

Instead of the dominant structural approach, Shepard (1987b) framed generalization as the reflection of a Bayesian inference problem: specifying the category $\boldsymbol{C}$ of stimuli that lead to a given consequence. In our example, $\boldsymbol{C}$ would be the set of colors that lead to the reward. **SHEPARD** assumed that the degree of generalization to a new stimulus $\boldsymbol{y}$ is proportional to $\boldsymbol{P(y \in C | X = x)}$, the probability that $\boldsymbol{y}$ belongs to the category $\boldsymbol{C}$, given the example $\boldsymbol{x}$. He then found sufficient conditions under which this function is approximately exponential. The ubiquity of the exponential law is ultimately explained by the fact that these conditions are reasonable for a wide variety of realistic problems. **T&G** extend Shepard's analysis to cases in which some of Shepard's conditions are not met: they allow multiple examples and nonconvex consequential regions. In both cases, violations of the exponential law are possible (see Fig. 2, 3, and 5 of **T&G**'s paper).

**SHEPARD**'s argument makes no distinction between conscious and automatic inferences. In addition, it is gloriously silent about representational and processing issues. In the functional approach, probabilities are just tools used by scientists to understand conditions under which observed behaviors are reasonable. These probabilities do not need to be explicitly represented by the organism under study. Consider, for example, the classic signal detection problem of discriminating a known signal in the presence of white noise. One can implement an optimal classifier for this problem without computing any probabilities at all. All the system needs to do is to estimate the correlation between the observed data and the known signal and to make decisions based on whether such correlation is larger than a threshold. Still, a Bayesian functional analysis in terms of subjective probabilities will be useful to

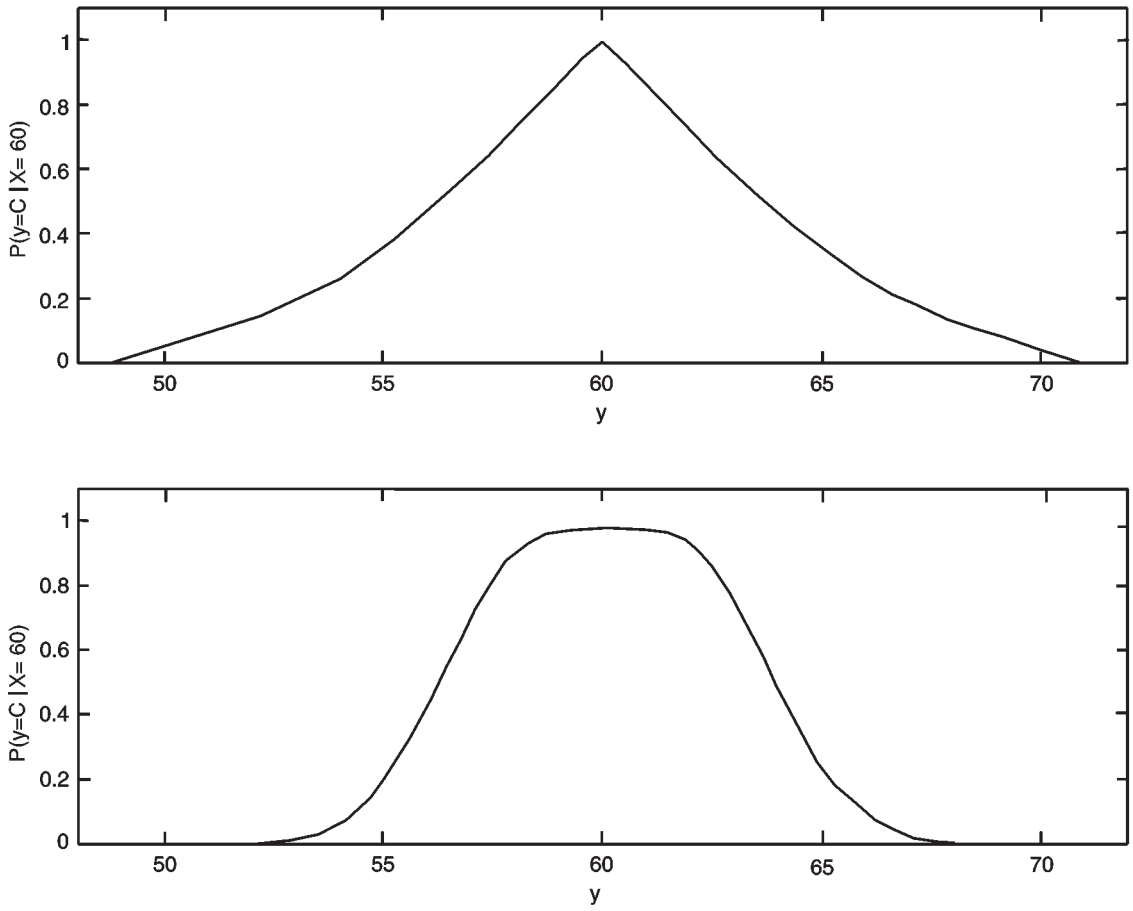



Figure 1 (Movellan & Nelson). Effects of non-uniform sampling on the generalization law. This figure shows the probability that a value $\boldsymbol{y}$ belongs to the interval $\boldsymbol{C}$, given the fact that the value 60 was sampled from that category. Top: The distribution of examples from $\boldsymbol{C}$ is uniform, resulting in a concave upwards generalization law. Bottom: The distribution of examples from $\boldsymbol{C}$ is Normal centered at $\boldsymbol{C}$ and with standard deviation equal to 1/8 of the length of $\boldsymbol{C}$. This results in a violation of the exponential law. In both cases the prior distribution for $\boldsymbol{C}$ was the same: uniform for location and positive truncated Gaussian for scale.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

understand the conditions under which the system is optimal.

The fact that we do not need to worry about how probabilities are represented makes the functional approach easily portable to a very wide variety of problems: from the study of single neurons, to the study of perception, conscious decision making, and the study of distributed cognition. It is thus not surprising that the emerging success of probabilistic functionalism reaches across a wide range of disciplines in the cognitive sciences: probability theory has become the language of choice to understand computation in neural networks (Bishop 1995). Bell and Sejnowski (1997) showed that the receptive fields of simple cells in primary visual cortex are optimal for transmission of natural images. Lewicki (2000) used similar techniques to show the optimality of cells in the auditory nerve. Knill and Richards (1996) illustrate the power of Bayesian techniques to understand perception. The "rational" movement in cognitive psychology is a perfect illustration of how functionalism can help us understand high level cognition (Oaksford & Chater 1998).

#### Extensions of the analysis of generalization.

Shepard (1987b) provided a solid foundation for the functional analysis of generalization and **T&G** extended the analysis in important ways. However, there are still some outstanding issues that need to be addressed. In this section we focus on such issues.

***Response rule.*** Consider the case in which an organism is rewarded for pecking a red key $\boldsymbol{x}$. The current analysis assumes that the rate of response to a novel key $\boldsymbol{y}$ is proportional to $\boldsymbol{P(y \in C \mid X = x)}$, the probability that $\boldsymbol{y}$ belongs to the consequential region $\boldsymbol{C}$, given the example $\boldsymbol{x}$. While this function exhibits the desired exponential law, it is unclear why one may in general want to respond with a rate proportional to such probability. For a functional explanation to be complete, this point needs to be addressed.

***Typicality and sampling.*** While **T&G** significantly generalize **SHEPARD**'s analysis, they still constrain their work to the following conditions: binary membership functions and uniform distribution of examples within categories. These assumptions may be still too restrictive. It is well known (Rosh 1978) that humans do not treat all elements of a concept equally (e.g., robins are better members of the category birds than penguins are) and, thus, graded membership functions may be needed to model human inference. Moreover, in many situations, examples do not distribute uniformly within categories and this may result in significant changes of the generalization function. For instance, take the case used by **T&G** of a doctor trying to determine the healthy levels of a hormone. A healthy patient has been examined and found to have a hormone level of 60. What is the probability that another hormone level, for example 75, is also healthy? In this case the consequential region $\boldsymbol{C}$ is an interval representing the set of healthy hormone levels. If we assume that hormone levels have a uniform distribution within that interval then the exponential law of generalization follows (see Fig. 1, top). However, if we let hormone levels to be normally distributed within the interval, that is, more probable about the center of the category than at the extremes, then the exponential law can be violated (Fig. 1, bottom).

***The size principle.*** According to the size principle proposed by **T&G**, smaller categories tend to receive higher probability than larger categories. Under the assumption that examples are uniformly sampled from categories, this is just a consequence of one of the axioms of probability. However, if examples distribute in a non-uniform manner within the category, for example, if robins are more likely to be sampled as members of the category "birds" than penguins are, the size principle would need to be reformulated, perhaps in terms of the entropy of the sampling distribution rather than the size of the category.

***Statistical analysis of the environment.*** **SHEPARD** and **T&G** frame their analyses in terms of subjective probabilities. Thus, it is entirely possible for the generalization law to be subjectively optimal and objectively inadequate. We believe a crucial part of probabilistic functionalism is to analyze the statistics of actual environments and to test whether the assumptions made by functional models are reasonable for the environments at hand. See Movellan and McClelland (2001) for an example of how this analysis may proceed in practice. This issue needs to be addressed in the context of the exponential law of generalization.

***Making predictions.*** Besides offering useful descriptive insights, probabilistic functionalism can also be predictive. For example, Movellan and McClelland (2001) analyzed a psychophysical regularity: the Morton-Massaro law of information integration. This law is observed in experiments in which subjects integrate two or more sources of information (e.g., the speech signal from a person talking and the visual information from the talker's lip movements). According to this law, ratios of response probabilities factorize into components selectively influenced by only one source (e.g., one component is affected by the acoustic source and another one by the visual source). Previous debates about this law centered on a structural issue: Is this law incompatible with interactive models of perception? Movellan and McClelland found that both feed-forward and interactive mechanisms can perfectly fit the law and thus this structural issue is undecidable. In contrast, a functional Bayesian analysis of the law helped find a novel task for which the Morton-Massaro law should be violated. Experiments confirmed this prediction. Similar predictive analyses would be helpful in the context of the generalization law.

## Beyond an occult kinematics of the mind

Keith K. Niall

*Defence and Civil Institute of Environmental Medicine, Defence Research and Development Canada, Toronto, Ontario M3M 3B9, Canada.*
**Keith.Niall@dciem.dnd.ca**

**Abstract:** The evidence for a kinematics of the mind is confounded by uncontrolled properties of pictures. Effects of illumination and of picture-plane geometry may underlie some evidence given for a process of mental rotation. Pictured rotation is confounded by picture similarity, gauged by gray-level correlations. An example is given involving the depicted rotation of Shepard-Metzler solids in depth.
**[HECHT; KUBOVY & EPSTEIN; SHEPARD; TODOROVIČ]**

Brave explorers have often mistaken the nature of the greatness they discover. Columbus thought he had sailed to the Indies; Frege thought he had reduced arithmetic to logic; **SHEPARD** thought he had found the kinematics of mind. SHEPARD proposes that the "abstract constraints of geometry" (i.e., the three-dimensional geometry of our terrestrial environment) are separable from the "concrete constraints of physics." He proposes that the former – essentially kinematic – constraints match what is represented in vision and visual imagination, better than other constraints which are characteristic of physical dynamics. That is, he draws a dichotomy between kinematics and dynamics, and bases his theory of representation on kinematics. The dilemma may be premature, for there is more to consider in a theory of vision and visual imagination. Conditions of illumination, and the perspective geometry of pictures should also be subsumed; for, though illumination and perspective geometry may not seem central to the psychology of representation, they are central to the study of vision.

Sometimes explanation is simpler or closer to hand than one may imagine. The "mental rotation" effect may not concern kinematics or rotation in three dimensions at all; I argue it depends mainly or wholly on the perspective geometry of pictures. The bulk of this commentary is devoted to presenting a small illustration of this point for some depicted rotations in depth.

Consider the Shepard-Metzler solid that is depicted in Figure 1. A single solid composed of many cubes has been rendered in six perspective views: that is, in six perspective pictures. These perspective pictures represent rotations in depth: rotations of −120%, −60%, 0%, 60%, 120%, and 180% about a vertical axis


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

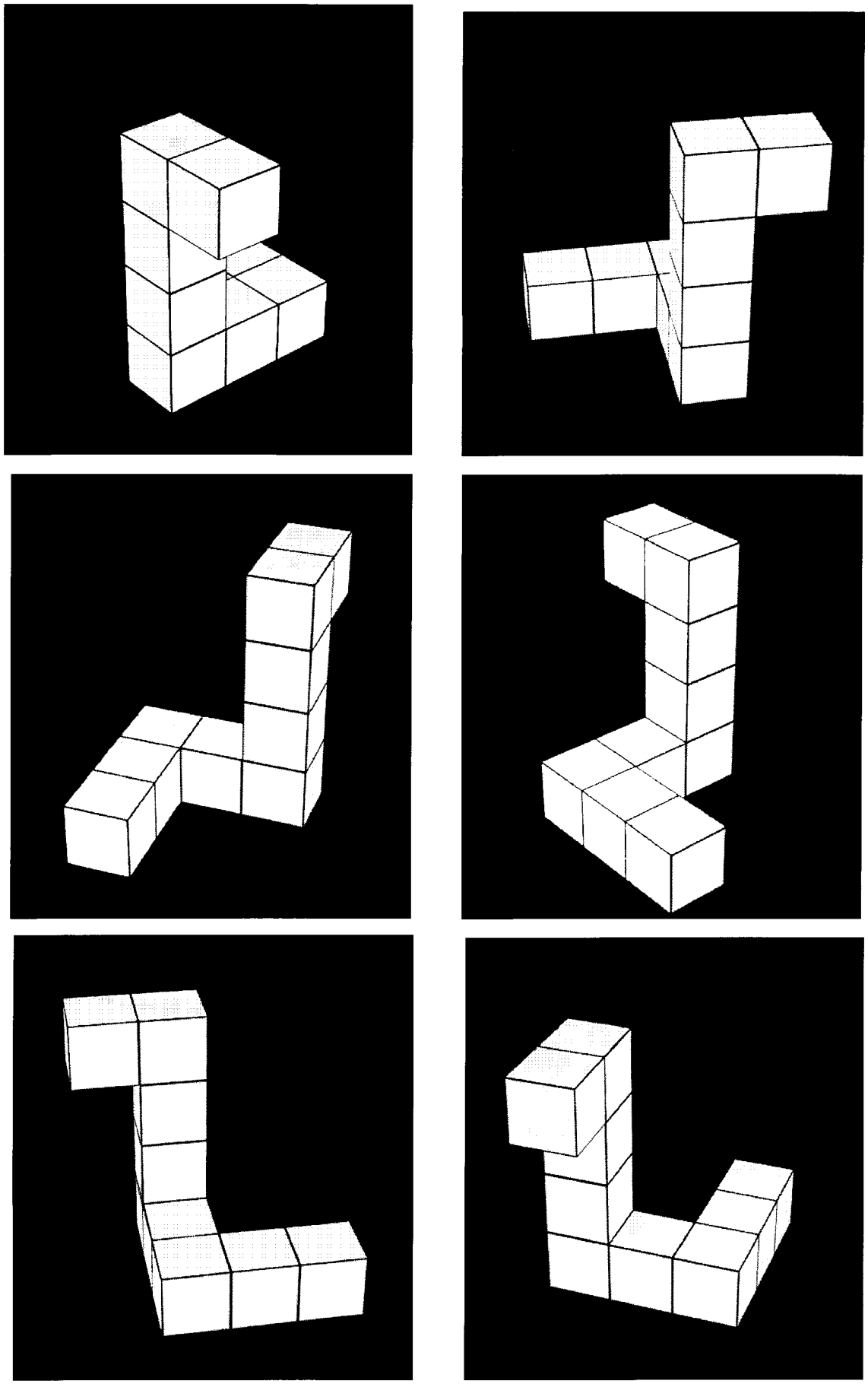


Figure 1 (Niall).



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

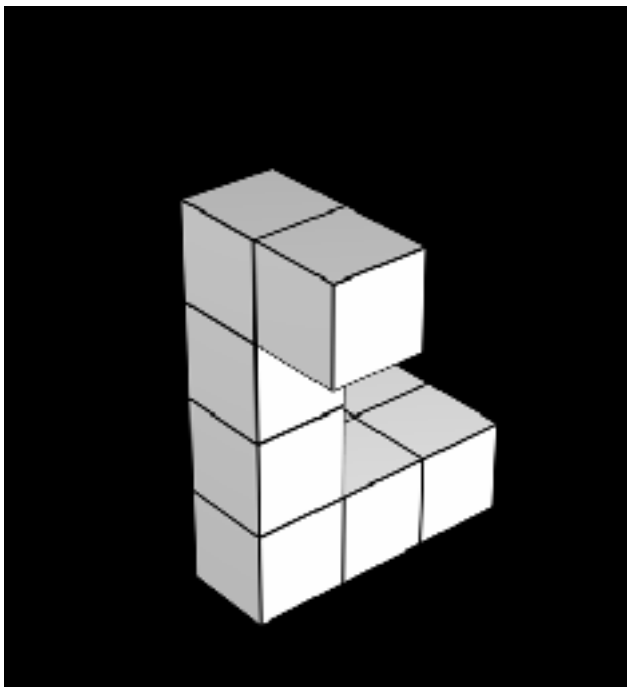


Figure 2 (Niall).

(following Fig. 1 in reading order). This solid is depicted as illuminated by one distant but strong light source, and also by two separate nearby but dim light sources.

Many perspective views of the solid can be generated: all the views separated by 20° intervals of rotation about a vertical axis were generated. A single view of another Shepard-Metzler solid is depicted in Figure 2. That three-dimensional solid is the mirror image (i.e., the *enantiomorph*) of the solid depicted in Figure 1.

Each perspective view or picture of the Shepard-Metzler solids can be considered as divided into a matrix of pixel elements. Every pixel is designated by a lightness value called a gray level, where a value of 0 stands for black and a value of 255 stands for white. (The original uncompressed images were 500 by 456 pixels in size.) The gray levels of two images of equal size can be correlated. Each image is represented by a matrix of numbers which stands for its gray levels. Then the correlation between two images is the correlation between their corresponding matrices of equal size. This operation of the correlation of gray levels enables a measure of the similarity of these images or pictures. The correlation of gray levels is a planar operation – it does not depend on an interpretation of the pictures as perspective views in depth. It is an elementary operation of image processing. Gray-level correlations were computed for many pairs of views of these two Shepard-Metzler solids. Some pairs included a view of the first solid (the middle picture of Fig. 1, left-hand side) as a standard picture. This standard picture was paired with a series of other views of the same solid. The depicted direction of rotation is clockwise as seen from above. Other pairs were formed by matching such views with the picture of the other Shepard-Metzler solid (Fig. 2). The depicted direction of that rotation is counterclockwise. (Note that angular difference is not well-defined for enantiomorphic solids.)

A gray-level correlations was computed on each pair. Again, the correlations are a measure of the similarity of the picture pairs in terms of gray levels. These correlations have been plotted in Figures 3 and 4 as a function of the depicted angular difference of the Shepard-Metzler solids. (The correlations are plotted as one minus the correlation value.) That is, these two graphs show the relation of depicted angular difference to the gray-level correlation

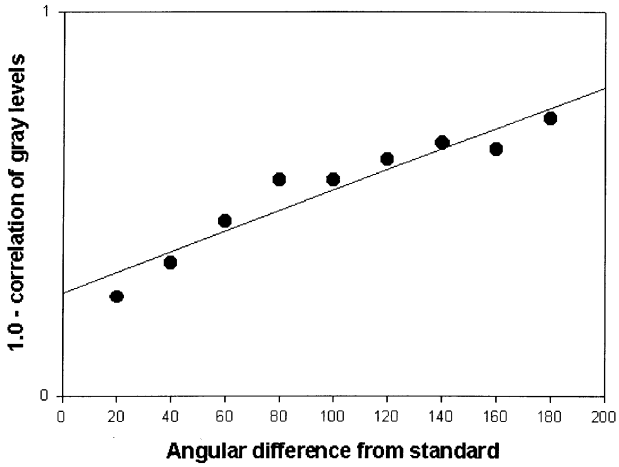



Figure 3 (Niall).

of the pictures. This correlational measure increases across pairs, as the standard view of the first solid is paired with other views of the first solid at 20° intervals from −20° to −180° inclusive (black dots, Fig. 3: the series includes the upper right and upper left images of Fig. 1). The measure also increases across pairs, as the mentioned view of the second solid (i.e., the "different" comparison, or "distractor" figure) is paired with views of the first solid at 20° intervals from 20° to 180° inclusive (white dots, Fig. 4). The series includes the middle right and lower left images of Figure 1; the picture pair of greatest similarity which includes Figure 2 is taken to mark 0° of angular difference. Many such series can be formed of views of either solid.

These trends bear strong resemblance to the response-time functions for the comparison of such picture pairs by observers – which ideal response-time functions are supposed to reflect the output of a mental operation of rotation in three dimensions. (The linear regression coefficient for the points of Fig. 3 is $r \approx 0.97$, and for those of Fig. 4 it is $r \approx 0.95$, each for nine points.) In other words, there is a correlation measure that can be computed in the picture plane, which may predict observers' response times to these picture pairs very well. Of course, *this* very simple correlation of gray levels is not likely to account for results on all the pictures which have been employed in manifold experiments on "mental rotation." But the pattern of these correlations indicates a broader possibility: it is the similarities of pictures, and not the kinematics of representations, that is key to understanding the

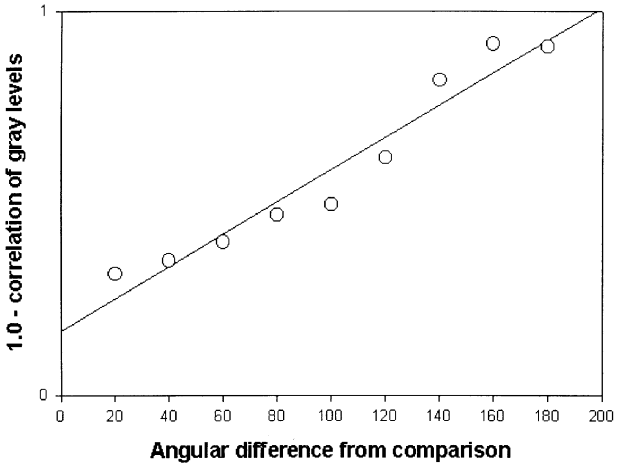



Figure 4 (Niall).


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

mental rotation effect. Other picture-plane correlations may be used in other cases, to describe response times and other data given as evidence for a mental process of rotation.

Where an operation of rotation was thought necessary to provide a criterion for judging the identity of solid forms, these correlations offer only a measure of similarity. Yet, such correlations might help explain why response times to these picture pairs are unique rather than bivalent (as an operation of rotation can proceed either the shorter way around, or the long way). They might explain why response times to views of solids paired with views of enantiomorphic (left- and right-handed) solids increase linearly with depicted angular difference. They might also explain why identical picture pairs are associated with substantially lower response times than picture pairs that represent a small rotation in space. These changes in correlation may seem a nuisance, a confounding variable in the search for a more complete characterization of the new mental kinematics. On the other hand, such measures on the surfaces of picture pairs could account for most of the story. Correlational measures (in contradistinction to invariants) may account for the mental rotation effect in depth without recourse to interpretation of the pictures as representations of depth (see also Niall 1997). Such an approach promises a simple, concrete account of some evidence in support of a kinematics of the mind.

The beauty of **SHEPARD**'s proposal for a kinematics of the mind is the dimly-reflected beauty of geometry, of invariants (i.e., the beauty of group structure and the invariant theory of classical kinematics). Yet we may not require that geometry of three dimensions to explain the experimental phenomena at hand. **HECHT** makes the astute claim that it is an *empirical* matter if invariants or other candidate regularities of the environment provide a model of some aspect of vision. HECHT makes another point that such invariants ought to be "non-trivial" – yet his and SHEPARD's examples are trivial ones which confuse invariants with recurrent environmental characteristics. Invariants are nothing like the direction of illumination for a standing observer, the conservation of water level, or else statistics over *unspecified* geometric properties. For a better account of invariants in the study of vision, see Mundy et al. 1994.

**TODOROVIČ** makes another strong point that our real knowledge of kinematics is based on a capacity for idealization, different from our ability to see. In contrast to Proffitt's suggestion to **KUBOVY & EPSTEIN**, one can say that motion *often* violates pure kinematics. The friction of rough surfaces, the surface tension of fluids, and many other physical effects underlie ordinary visible phenomena, but do not enter into the idealizations of kinematics. Also, the development of knowledge about kinematics itself is not a story of internalization: such a claim would fictionalize the history of science. The development of kinematics is not a chapter in a psychology of the individual, since the development of physics has supposed an epistemological division of labor. It is not a chapter in the evolutionary psychology of the species either, since organisms adapt to existing local conditions, and not to counterfactual or universal conditions. A psychology which fails to acknowledge the place of ideals in its description of intellectual competence – including competence in kinematics – is a psychology which fails to draw a cogent distinction between perception and thought.

The notion of an internalized kinematics addresses a fundamental problem in psychology – better said, the notion scratches a deep conceptual itch. **HECHT** claims that the notion solves one of the hardest problems in the study of perception, the underspecification problem. **KUBOVY & EPSTEIN** describe the inverse projection problem as fundamental to the problem of vision; the inverse projection problem revisits the underspecification problem for vision. This problem is neither deep nor hard nor fundamental; its conceptual itch is illusory – if anything, the problem is the result of a deep confusion. (Kubovy & Epstein cite James Gibson (1979) as calling it a "pseudoproblem.") No solution involving an internalized kinematics is required where there exists no problem.

Some psychological phenomena like "mental rotation" may arise as a consequence of the characteristics of illumination, or the perspective geometry of pictures. **SHEPARD** remarks that the evolutionary significance of the invariant characteristics of light-reflecting objects is primary to that of the characteristics of light or light sources. Yet for vision, the invariants preserved and the variants generated in the propagation of light are primary to other "invariants" of light-reflecting objects – those which are not preserved when reflected light reaches our eyes. Of course we might prefer to expound the psychology of representation without any detour of discourse about the senses, including the sense of sight. But "in the acutal use of expressions we make detours, we go by sideroads. We see the straight highway before us, but of course we cannot use it, because it is permanently closed" (Wittgenstein 1953/1967, p. 127e).

## Functional resemblance and the internalization of rules

Gerard O'Brien and Jon Opie

*Department of Philosophy, University of Adelaide, South Australia 5005 Australia.* **{gerard.obrien; jon.opie}@adelaide.edu.au**
**http://www.arts.adelaide.edu.au/Philosophy/gobrien.html**
**http://www.arts.adelaide.edu.au/Philosophy/jopie.html**

**Abstract:** Kubovy and Epstein distinguish between systems that *follow* rules, and those that merely *instantiate* them. They regard compliance with the principles of kinematic geometry in apparent motion as a case of instantiation. There is, however, some reason to believe that the human visual system internalizes the principles of kinematic geometry, even if it does not explicitly represent them. We offer functional resemblance as a criterion for internal representation.
**[KUBOVY & EPSTEIN]**

According to **KUBOVY & EPSTEIN** (K&E), there are two ways of construing the fact that the perceived paths of apparently moving objects conform to the principles of kinematic geometry (Shepard 1994, pp. 4–6). One might suppose, with **SHEPARD**, that our visual system proceeds by applying internal knowledge of kinematic geometry. Alternatively, one might suppose, as K&E urge, that our visual system proceeds *as if* it possessed knowledge of kinematic geometry. The latter is always an option, say K&E, because of the difference between physical devices that *follow* rules and those that merely *instantiate* them (see target article, p. 619). Although K&E don't elaborate, their supporting discussion suggests they have in mind the well known distinction between physical systems whose behaviour is driven by internally represented rules (such as stored program digital computers) and those whose behaviour merely conforms to rules/laws, without internally representing them (the approximate conformity of the planets to Newton's universal law of gravitation is the standard example). There is, however, some reason to believe that the human visual system does not merely instantiate the principles of kinematic geometry. Consequently, if the visual system does behave in accordance with these principles, as K&E concede, it must internally represent them in some way. We will argue for this view by briefly re-examining the distinction between rule-following and rule instantiation.

The paradigm case of a device whose behaviour is driven by represented rules – of rule-following – is the Turing machine. The causal operation of a Turing machine is entirely determined by the tokens written on the machine's tape together with the configuration of the machine's read/write head. One of the startling features of a Turing machine is that the machine's tape can be used not only to store data to be manipulated, but also to *explicitly* represent the computational rules according to which this manipulation occurs. This is the basis of stored program digital computers and the possibility of a Universal Turing machine (one which can emulate the behaviour of any other Turing machine).

This neat picture gets a little messy, however, when we consider that not all of the computational rules that drive the behaviour of a Turing machine can be explicitly represented in the form of to-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

kens written on the machine's tape. At the very least, there must be some primitive rules or instructions built into the system in a nonexplicit fashion, these residing in the machine's read/write head. Since these "hardwired" rules are not encoded in the form of discrete tokens written on the machine's tape, many theorists claim that they are *tacitly* represented (see, e.g., Cummins 1986; Dennett 1982; O'Brien & Opie 1999; Pylyshyn 1984). But what licenses this terminology? Is there any real difference between the behaviour of a Turing machine driven by "tacitly represented" rules and a planet obeying Newton's laws?

We think there is. Consider a Turing machine that adds integers. Such a machine receives as input a set of tokens representing the numbers to be added, and eventually produces further tokens representing their sum. Since the sequence of tokens on the machine's tape (representing both summands and sums) is a set of discrete physical objects, the Turing machine's operation can be characterised in terms of a pattern of causal relations among its tokens. From this perspective, the Turing machine succeeds in adding numbers because the causal relations among its inputs and outputs, considered as physical objects, mirror the numerical relations among sums and summands. The computational power of the Turing machine thus depends on the existence of a homomorphism between an empirical relational structure (in this case a causal one) and a mathematical relational structure, as **K&E** would put it (sect. 2.1, p. 621). We will characterise the relationship between the system of tape tokens and the integers as one of *functional resemblance.* One system functionally resembles another when the pattern of causal relations among the objects in the first system preserves or mirrors at least some of the relations among the objects in the second (for further discussion see O'Brien & Opie, forthcoming).

If, by virtue of the causal relations among its internal states, a physical system functionally resembles some domain D, then in our view it is appropriate to interpret the mechanisms that drive the system as *internal representations* of the relations between the objects in D. In the case of our imagined Turing machine, D is the (abstract) domain of integers, which are subject to various arithmetic relations, including those codified in the rules of addition. Consequently, it's appropriate to interpret the Turing machine as embodying internally represented rules of numerical addition. It does not matter whether the Turing machine explicitly represents these rules in the form of tokens written on its tape, or tacitly represents them courtesy of the configuration of its read/write head. What matters is that the casual relations among some of its internal states mirror specific mathematical relations among the integers.

Functional resemblance serves to distinguish devices like the Turing machine, which represent rules, from other physical systems that merely conform to rules. In the case of the solar system, for example, while the motions of the planets respect Kepler's laws, which can in turn be derived from Newton's universal law of gravitation, there is little sense to be made of the idea that these laws are internally represented by the system. Such laws are actually *our* attempt to represent (in mathematically tractable form) the regularities inherent in the causal dynamics of the system. Thus, when we simulate the planetary motions on a digital computer, we arrange things so that the causal relations among some of the internal states of the computer mirror the geometric and dynamical relations among the planets. The functional resemblance runs from simulation to planetary system, not the other way around. We are thereby warranted in saying that the inherent gravitational constraints of the solar system are *represented* in the computer, but *not* that the solar system represents the laws of motion – it merely instantiates them (to use **K&E**'s language).

What, then, of the human visual system's conformity to the principles of kinematic geometry, at least where the behaviour of apparently moving objects is concerned? Here it would seem that a relationship of functional resemblance does obtain between internal states of the human visual system and the motions of objects in the world. Of course, we don't yet know which brain processes are responsible for producing experiences of apparent motion. But it is reasonable to infer that the causal processes involved are systematically related to the structure of the experiences themselves. Such experiences portray objects that are subject to the kinds of constraints identified by Shepard, namely, they are conserved, are restricted to movements in two or three dimensions, and traverse kinematically simple paths (Shepard 1994, pp. 4–6). By assumption, these constraints are mirrored in the causal relations among the neural vehicles of apparent motion: there is a functional resemblance (if not an isomorphism) between brain states and perceived paths. Although real objects do not invariably move in accordance with kinematic constraints, the motions delimited by those constraints certainly constitute a class of *possible* object motions. Indeed, motions defined with respect to axes of symmetry are common in the context of manual object manipulation. By the transitivity of resemblance, we thus establish that there is a functional resemblance between the system of internal vehicles responsible for experiences of apparent motion and the motions of real objects acting under kinematic constraints. In light of our earlier discussion, this suggests that we may regard the visual system as representing the principles of kinematic geometry, not merely instantiating them.

Even if the principles of kinematic geometry are not explicitly encoded by the visual system, it therefore appears that kinematic principles *are* "lodged in the mind" (**K&E**, p. 619). Kinematic constraints are built into the very fabric of the visual system. They are not merely "passive guarantors or underwriters that are external to the perceptual process," but "active constituents in the perceptual process" (ibid.) (at least under the stimulus conditions that give rise to apparent motion). In other words, we must reject **K&E**'s modest interpretation of **SHEPARD**'s observations, leaving Shepard's own conclusion: kinematic constraints are internally represented, because they have been "internalized" by the brain.

## The mathematics of symmetry does not provide an appropriate model for the human understanding of elementary motions

John R. Pani

*Department of Psychological and Brain Sciences, University of Louisville, Louisville, KY 40292.* **jrpani@louisville.edu**
**http://www.louisville.edu/~jrpani01/**

**Abstract:** Shepard's article presents an impressive application of the mathematics of symmetry to the understanding of motion. However, there are basic psychological phenomena that the model does not handle well. These include the importance of the orientations of rotational motions to salient reference systems for the understanding of the motions. An alternative model of the understanding of rotations is sketched.
**[SHEPARD]**

In even the most elementary domains of physical understanding, there are clear distinctions between problems that are natural and intuitive for people, and ones that are challenging. These phenomena extend into many areas of cognition, including spatial organization, object recognition, and event knowledge; and explanation of this variation in physical understanding is an important undertaking for cognitive theory.

In the first part of his article, **SHEPARD** constructs an explanation of variation in our understanding of elementary motion from the modern mathematics of symmetry. In this view, our understanding of motion is embodied in a six-dimensional manifold, and those motions that are natural for us to perceive or imagine are the geodesics in the manifold; the structure of the manifold, and the lengths of the geodesics in it, are reduced when the objects that move are rotationally symmetric. This geometric model is an impressive achievement, and it appears to represent a thorough exploration of the application of this mathematics to spatial cognition. Despite containing important elements of truth, however, I


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

think the model is insufficient. In the remainder of these comments, I confine my discussion to the case of pure rotation, and I draw upon relatively recent studies that demonstrate breakdowns in the perception or visualization of rotations (though not of continuous motion; e.g., Massironi & Luccio 1989; Pani 1989; 1993; Pani & Dupree 1994; Pani et al. 1995; Parsons 1995).

It is often much easier for a person to perceive or visualize a rotation if the object has an intrinsic axis aligned with the axis of rotation. In terms of a standard manifold, some geodesics are preferred to others. To account for this finding, **SHEPARD** equates intrinsic object axes with axes of rotational symmetry, and the manifold shrinks in accordance with the symmetry. However, it is not necessarily object symmetry that determines an object axis (e.g., Pani et al. 1995, Experiment 2). In response, SHEPARD appeals to local symmetry in objects – symmetry of object parts – but it is unclear how this development improves the fit of the model to the phenomena. It is the whole object that rotates, and the manifold presumably takes account of the whole object. In addition, the appeal to local symmetry is based largely on intuition. The central segment of the Shepard/Metzler figure (see the target article) has a fourfold symmetry, but perhaps this segment is a salient object axis because its global shape approximates a long central line (a main axis).

Determining intrinsic axes of an object is analogous to finding the best fit of an equation to a complex set of data. A variety of variables are important, no single variable, such as symmetry, is necessary, and the fit may be approximate. If object axes are not clearly due to rotational symmetry axes, **SHEPARD**'s model is not clearly applicable to the issue of object structure in the understanding of rotations.

A second problem for the proposed model is that people understand certain rotations much better when the axes of rotation are vertical (or in some cases Cartesian; Pani et al. 1995). This powerful effect of the primary axis of the environmental reference system should be explained by a theory of the understanding of rotations, but it appears to have no special place in **SHEPARD**'s model.

More generally, the model appears not to take account of the varying salience of alternative reference systems in the understanding of orientation and rotation (e.g., the vertical enclosures, preferred object structures). However, basic phenomena in this area pertain to the presence of multiple reference systems and the resulting definitions of orientation (see below).

It is possible to construct alternative accounts of the understanding of rotation that handle more of the basic phenomena and that are equally consistent with notions of biological adaptation. Consider the main points of a fairly simple model (see also Pani 1997; 1999):

1. Orientation of an object is determined relative to a reference system.

2. There are three psychologically real dimensions of orientation relative to a reference system. In the terrestrial environment, these are slant to the vertical axis, radial direction in the horizontal plane (e.g., compass direction), and spin of an object about its own axis.

3. Where multidimensional descriptions are used, one parameter (i.e., one-dimensional) variation is psychologically simple.

4. People understand rotations as one-parameter continuous changes of orientation.

5. Although the kinematic geometry of a rotation defines a possible reference system with which the motion is self-aligned, this system is not generally salient for people. Hence, a rotation is not seen as a one-parameter change of orientation when the motion and its geometry are the only reference system for the determination of that orientation.

6. The vertical of the environment and the intrinsic axes of objects are effective spatial reference systems for the determination of orientation.

When a rotation takes place about the vertical, it is readily seen as a one-parameter change of orientation, because the change takes place about an axis that is used to determine object orientation. Similarly, when a rotation takes place about an object axis, it is a one-parameter change of the object's perceived spin. (There is a similar description if the motion is described in an object-centered reference system.)

When actual or potential rotations are not aligned with objects or with the vertical, they can be very difficult to perceive or imagine. The orientation of the object is defined in terms of the relations between the object axes and the environment, but the rotation takes place around an axis that is not aligned with either of those reference systems. In perception of such rotations, the object appears to go through a more or less incoherent, but continuous, change of orientation (Pani et al. 1995). Relative to the reference systems used to perceive the orientation of the object, that is indeed what it is happening. In a case such as this, the unperceived rotation is a geodesic in some kinematic space, but not in the reference system that is salient for the perceiver.

## Evaluating spatial transformation procedures as universals

Lawrence M. Parsons

*Research Imaging Center, University of Texas Health Science Center at San Antonio, San Antonio, TX 78284.* **parsons@uthscsa.edu**

**Abstract:** Shepard proposes that the human mind relies on screw displacement because of its adaptive simplicity and uniqueness. I discuss this hypothesis by assessing screw displacement with respect to (1) other plausible spatial transformations, (2) a variety of criteria for adaptive efficiency and utility, and (3) a variety of psychological conditions in which observed responses discriminate amongst alternative spatial procedures. **[SHEPARD]**

The proposal of screw displacement as a perceptual-cognitive universal (Shepard 1984; and target article) is illuminated by being considered in a broad context. First, for purposes of comparison, it is useful to compare screw displacement to two other basic procedures that use straight-line translation between initial and final positions in conjunction with the following simultaneous and homogeneous reorientation. The "shortest trajectory" reorientation produces a rotation about an axis unique for each orientation difference that minimizes rotation. The "spin-precession" reorientation (Parsons 1987a; 1987b; 1987c) produces a rotation about an instantaneously changing axis produced by simultaneous rotations about two orthogonal axes (e.g., a principal axis of the object and an axis of the environment, as in the precession of a spinning top).

Second, it is useful to examine the three spatial procedures with respect to a variety of criteria for efficiency and utility, such as the following: (1) To produce a trajectory for any pair of initial and final orientations, or to produce any arbitrary trajectory. (2) To produce a trajectory that minimizes translation and rotation. (3) To possess efficient computations for trajectory planning and execution. (4) To produce trajectories that are predictable. (5) To produce trajectories in which the displaced object's position and velocity are smooth functions of time.

Each procedure above possesses the first property but they differ greatly in satisfying the other criteria. With respect to minimizing trajectory over all possible initial-final orientation/positions, the "shortest trajectory" is perfectly efficient in the Euclidean metric $(R)_3$, followed in efficiency by the "spin-precession," which is more efficient than screw displacement. Screw displacement is increasingly inefficient as the difference in initial and final position grows beyond the size of the object. With respect to a bi-invariant Riemannian metric on $SO_{(3)}$, screw displacement minimizes distance by using a single operator (rather than separate rotations and translations), but it is inefficient for many spatial cognition tasks in which Euclidean displacement is proportional to arduous mental simulation time.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

It is difficult to assess the efficiency of computations for planning and execution without greater knowledge of the representations, implementations, and systems involved (Marr 1982). At present, this property may be best assessed by evaluating the capacity and efficiency of human performance. The evidence for the use of screw displacement in apparent motion seems weak. Cases (a) through (i) in Figure 2 of Shepard's (1984) paper are all special cases in which each procedure here would produce identical trajectories. In cases (j)–(l), screw displacement is so slightly different from the trajectories produced by the other procedures that an extraordinarily sensitive measure would be required to show that screw displacement was the perceived trajectory. The orientation and position differences used in the Bundeson et al. (1983) study, also cited by Shepard (1984), are all special cases in which the procedures here would produce identical trajectories. In Foster's (1975b) experiment, measures were taken of the perceived intermediate orientation and vertical position of an apparently moving rectangle for two subjects. In none of eight distinct differences in position and orientation did the evidence on both measures favor screw displacement. In four cases, measures for vertical orientation were consistent with screw displacement, but the orientation measures were inconsistent with both screw displacement and straight-line translation with simultaneous rotation. In the other two cases, measures for vertical position were consistent with screw displacement but measures for orientation were consistent with straight-line translation superimposed on rotation. There was no evidence for the use of screw displacement in two later, related studies (Farrell 1983; Mori 1982).

Recent research on aspects of spatial perception and cognition other than apparent motion indicates humans do not readily conceive of object orientations in terms of shortest path axis and angle. In one study (Parsons 1995), individuals of high spatial ability were shown in most cases to be unable to imagine an object rotate about an axis and angle so as to accurately envision its appearance. Nor could they conceive of the axis and angle by which the object would rotate in a shortest path between two orientations. Human accuracy improves when there are special spatial relationships among the object's parts, the rotation axis, and significant directions in the viewer-environment frame (for related findings, see Pani 1993). Such data suggest that, in general, humans apparently cannot readily conceive of the minimum angle rotation, which is at the heart of the computations required for screw displacement and shortest trajectory.

The fourth property, the ability of a procedure to produce predictable trajectories, is adaptive because the system can determine whether a trajectory is suitable with less than the full computation of a trajectory, thus shortening initial planning. Spin-precession produces the most readily predictable path. By comparison, screw displacement is very difficult to predict when there are both position and orientation differences: an object could pass through any region of space between initial and final positions, depending on the difference in orientation. There is no obvious rule-of-thumb to predict screw displacement in order for the trajectory to be checked for obstructions without computing and examining the full path.

The fifth property, the ability of spatial procedures to produce smooth trajectories, is useful partly because execution time may increase for extra processing required to produce changes in velocity. Each of the three procedures produces smooth trajectories, however, they differ in how gracefully they degrade with unexpected online changes in target orientation. Straight-line translation with simultaneous reorientation has the advantage because differences in position and orientation are eliminated by independent "operators" and the effect of changing the goal orientation mid-trajectory has a relatively minor effect on the trajectory. Screw displacement requires a whole new computation and a potentially radical change in trajectory (depending on the difference in the original and new goal orientation).

In summary, with the data currently at my disposal, I believe it is premature to conclude that the screw displacement is a universal for the human mind. This matter can be clarified by additional studies, particularly those varying stimulus conditions to discriminate amongst various alternatives, by greater analytical considerations of broad functional utilities of various spatial transformations and by a reconciliation of spatial procedures implemented across perceptual-cognitive processes.

## Reflections on what timescale?

John Pickering

*Psychology Department, Warwick University, Coventry, CV4 7AL United Kingdom.* **j.pickering@warwick.ac.uk**
**http://www.warwick.ac.uk/staff/J.Pickering**

**Abstract:** Recent developments in both evolutionary theory and in our ideas about development suggest that genetic assimilation of environmental regularities may occur on shorter timescales than those considered by Shepard. The nervous system is more plastic and for longer periods than previously thought. Hence, the internal basis of cognitive-perceptual skills is likely to blend ontogenetic and phylogenetic learning. This blend is made more rich and interactive by the special cultural scaffolding that surrounds human development. This being so, the regularities of the environment which have been genetically assimilated during the emergence of modern human beings may themselves be the products of human action.

**[SHEPARD]**

If there has been the "Evolution of a mesh between principles of the mind and regularities of the world" as Roger **SHEPARD** put it (1987a), then we might ask what "regularities" are in question and what timescale is appropriate for this evolution. The regularities dealt with in most of his work reflect the structure of space-time, the properties of energy spectra, and the fact that objects having the same behavioural significance to an organism also have a tendency to resemble each other. Now indeed we might expect that genetic influences on the perceptual-cognitive abilities of organisms would reflect such regularities. After all, they are crucial features of the environment in which organisms have learned to survive and reproduce. In fact, our present ideas about evolution make the proposal almost tautologous.

But these ideas are changing. Moreover, they are changing in a way that make it possible to extend **SHEPARD**'s approach beyond regularities which are either on very long time scales or, since they are universal properties of time, space, and energy, do not have any timescale at all. Now, conditions local in space and time to terrestrial life, such as the period of the earth's rotation, are actually not universal but contingent. However, when considering the genetic assimilation of the regularities of the world, they have been in place for sufficiently long enough to be treated as if they were. Thus, such long-term contingencies may well have an influence on genetic assimilation that is effectively equivalent to universal regularities which are actually necessary rather than contingent.

But once this step is taken, we may ask just how long "long-term" actually is and the complementary question of how quickly genetic change occurs. This is where recent work in evolutionary theory comes in. This, generally speaking, has proposed that a broader range of factors underlie the variation on which selection acts. These factors include development, learning, and emergent interactions within the organism-environment system (see e.g., Van de Vijver et al. 1998 for a review). Now, such factors are active over far shorter time scales than the factors hitherto seen as the primary sources of variation, such as mutation, recombination, and genetic drift in isolated populations. This means that what we might call the "responsiveness" of the genotype, is actually a rather more lively and flexible matter than the standard neo-Darwinism account of evolutionary change suggests.

Phylogenetic change seen in this light can be considered as intimately bound up with ontogenetic change and both can be seen as forms of learning, albeit taking place over different time scales


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

(e.g., Hinton & Nowlan 1996). This is particularly easy to do when we consider the open end of the open *versus* closed continuum of evolutionary strategies, an idea originally proposed by Ernst Mayr (1976) and developed by Karl Popper (1978). Closed strategies rely on instinctive patterns of perception and action. These are immediately available as the animal develops, although they commit an organism to a particular niche. Open strategies are developmentally more costly and depend on learning but make the animal much more adaptable. As uniquely active, flexible, and culturally shaped organisms, human beings occupy an extreme position at, or even beyond, the open end of this continuum.

This idea is supported by other recent changes in our ideas about development which complement the changes in evolutionary theory mentioned above. These, broadly speaking, treat development within the framework of dynamic systems theory (e.g., Dent-Read & Zuckow-Golding 1997) and accordingly make it more parsimonious to treat developmental change and evolutionary change as continuous (e.g., Butterworth et al. 1985). Likewise, the functional architecture of the brain itself is now treated as more open to environmental influences during development (e.g., Edelman & Tononi 2000; Elman et al. 1996). Treating evolution, learning, and development together in this way makes it seem likely that the evolutionary history of the human species has been marked by the genetic assimilation of regularities on far shorter timescales than those yet considered by **SHEPARD**. This has been a factor in the revival of interest in the Baldwin Effect – the idea that learning influences both the speed and the direction of evolutionary change (see e.g., Depew 2000).

The cultural scaffolding which surrounds and supports human development is so uniquely powerful a determinant of our phenotype that it is likely to have exerted a significant effect on the evolutionary emergence of modern human beings. Recent work has shown how the accumulation of the products of human cultural activity has influenced evolution (e.g., Deacon 1997) and has done so to such an extent that the very material structure of the environment to which human beings adapt, both by learning and by evolutionary change, should properly be seen as a social product (e.g., Ingold 1996). Putting these ideas together leads to the proposition put forward by Kingdon that: "Humans have become intrinsically different from apes by becoming, in a limited but very real sense, artefacts of their own artefacts." (Kingdon 1993, Ch. 1). That is, the environment which modern human beings have genetically assimilated is itself a human product. Thus, if the human genome does code for cognitive-perceptual skills, the regularities they reflect may well have appeared on shorter timescales than those hitherto dealt with by **SHEPARD**.

## Context effects equally applicable in generalization and similarity

Emmanuel M. Pothos

*Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, United Kingdom.* **e.pothos@bangor.ac.uk**
**http://www.bangor.ac.uk/~pss41b/**

**Abstract:** Shepard's theoretical analysis of generalization is assumed to enable an objective measure of the relation between objects, an assumption taken on board by Tenenbaum & Griffiths. I argue that context effects apply to generalization in the same way as they apply to similarity. Thus, the need to extend Shepard's formalism in a way that incorporates context effects should be acknowledged.
**[SHEPARD; TENENBAUM & GRIFFITHS]**

**TENENBAUM & GRIFFITHS** (**T&G**) note that **SHEPARD**'s formalism derives its elegance partly from its generality: the generalization function is supposed to be universal, with virtually no conditions placed on applicability. However, this universality seems to imply an objective measure of the relation between objects. This view appears to be shared by T&G who state that "generalization can be stated objectively . . . , while the question of how similar two objects are is notoriously slippery and underdetermined." When discussing the term p(h), T&G appear to acknowledge that context might influence generalization; however, there is no indication of how this would affect SHEPARD's formalism.

I would like to argue that generalization relations are subject to context effects in the same way as similarity ones are. I start with a brief illustration of context effects in similarity, then discuss how such effects also influence generalization.

***Context effects in similarity.*** Murphy and Medin (1985) argued that people's naive theories make objects cohere together into categories; they criticized a conception of category coherence based on similarity. However, it is possible that naive theories simply modify the perceived similarity of a set of objects, until these objects become similar enough to form a category. This would correspond to an interpretation of category coherence as grounded on similarity; however, this new similarity would be so susceptible to background knowledge and theories (for our purposes, the same as "context") that it would make it compatible with Murphy and Medin's approach. Barsalou (e.g., 1985) is more direct on this issue: if your house is on fire, then your university degree, the video player, and your passport are all objects that make up a perfectly acceptable category. In other words, this particular context will make the above objects similar enough for them to cohere into a category.

Goodman's (1972) concerns with similarity are reflected in studies like the above (Barsalou 1985; Murphy & Medin 1985): any two objects are potentially infinitely similar as they are consistent with each other along a potentially infinite number of dimensions. For example, a pen and an elephant are heavier than one milli-gram, two milligrams, and so forth. If we were to view this through the filter of Barsalou's work, we could say that it is always possible to provide a context that would make a set of objects as similar as we like.

Failing to capture context effects in a principled way has made many psychologists distrust similarity, a point made by **T&G** as well (see, e.g., Goldstone 1993; Hahn & Chater 1997). But does the flexibility of similarity implicate similarity's lack of utility in psychology? We shall return to this shortly.

***Context effects in generalization.*** Considerations analogous to the above apply in establishing that generalization is affected by context.

Generalization is a judgment of whether two objects can be considered to be equivalent. Consider a forest where different berries can be found. The most obvious variation in these berries is color. Let's suppose color can be represented as a single real number along a scale of 0 to 100. A person has tried, say, a berry whose color corresponds to a value of 60, and she has found this berry extremely tasty. What other berry colors should she expect to correspond to tasty berries? This is a typical scenario within which **SHEPARD**'s analysis is motivated. But say we change the last part of the scenario to: "and she has found that this berry relieves headaches. What other berry colors should she expect to correspond to berries that relieve headaches?" In the first case we require a judgment of equivalence on the basis of "taste" and in the second on the basis of "relieving headaches." In fact, there are infinite possibilities for judging equivalence between two objects. Under SHEPARD's formalism, each object is represented as a unique point in a psychological space, and the degree of equivalence of two objects is an objective function of the distance between them. Thus, all possibilities for equivalence are reduced to a single judgment. However, as is the case with similarity, whether two objects are equivalent or not will depend on the context we are interested in: two objects may be equivalent as far as taste goes, but not in terms of their health properties. Similar examples can be thought of with simpler stimuli, whose confusability may be affected depending on the particular background in which they are presented. Using the Figure 6 example in **T&G**, the arrangements of shapes can be presented along stimuli that emphasize relational versus primitive features; such differences in emphasis are likely to affect generalization.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

So we can ask whether different objects are equivalent or not, for different purposes, like "tasty food" or "substances affecting health." One could equate such purposes with **SHEPARD**'s consequential regions, in which case it might seem that Shepard's formalism does take into account the above considerations – the averaging process via which the exponential similarity function is derived assumes that two objects are members of several different consequential regions. But in practice, in any given situation, we will try to assess whether two objects are equivalent or not only for a particular purpose.

***Summary***. Context dependence appears to be important both in similarity and generalization. Context provides information that enables an observer to focus on the parts of the objects that are immediately relevant (see the introductory pages of **T&G**'s article). Without such information, the representation of an object could be anything, as Goodman's (1972) arguments compellingly illustrate, as well as the extensive literature on unsuccessful attempts to pinpoint the essential elements or "essences" of the concepts we have (e.g., Malt 1994; Medin & Ortony 1989; Pothos & Hahn 2000). Of course, how context is computationally accommodated within existing models of similarity or generalization is an open question (e.g., Heit 1997b; Nosofsky 1989; Pickering & Chater 1995; Tversky 1977); this commentary only extends as far as arguing for the need to acknowledge the relevance and importance of context in generalization as well.

## Shepard's pie: The other half

Karl H. Pribram

*Department of Psychology, Georgetown University, Washington, D.C. 20057.*
**pribramk@georgetown.edu**

**Abstract:** Having seen the development of Shepard's program at close hand, I have been inspired by the sophistication of his results. However, his program deals with only half of what is needed: Shepard's research tells *what* the perception/cognitive process is about, it does not tell *how* that process is implemented. True, Shepard has recourse to the "how" of process in evolution, but that is not the "how" of everyday implementation. For that we need to know the *brain processes* with which we can implement Shepard's insights.
**[SHEPARD]**

Roger **SHEPARD**'s "Perceptual-Cognitive Universals as Reflections of the World" is brilliant in conception, in the long series of experiments that provide substance to the conception, and in the lucid, succinct presentation in the review presented here. The key to understanding perception and cognition is, according to SHEPARD, the creation, by way of evolution, of representations of universals, invariances consisting of worldly properties tuned to survival. Shepard details his experimental results with regard to: (1) the perception of material objects; (2) color; and (3) categorizing according to objects of the same kind.

A good deal is known about the brain processes which implement **SHEPARD**'s program. With regard to the perception of objects, Eloise Carlton and I developed a program of research which integrated brain mechanism with Shepard's three-dimensional Euclidian character of physical space (Pribram & Carlton 1987; reviewed by Pribram 1991, in Lecture 5). With respect to color, DeValois and DeValois (1993) have provided a superbly detailed psychophysical/neurological process that accounts for issues such as the paucity of foveal cones sensitive to the higher (blue) end of the visible spectrum. Finally, there are initial attempts to describe brain mechanisms that categorize objects of the same basic kind. These programs of research include those by Martha Wilson (1987; Wilson & Debauche 1981) as well as my own work (Pribram 1986; reviewed by Pribram 1991, Lecture 10). Mathematical implementations have been developed in a collaboration between myself, Kunio Yasue, and Mari Jibu (Pribram 1991, Appendices F & G). Edelman's (1989) program is also relevant.

I was alerted to the possibility that group theory could account for object perception by William Hoffman (1966) and discovered that Lie actually invented continuous groups to account for object perception in a correspondence that took place at the end of the nineteenth century between Poincaré, Helmholtz, and Lie (reviewed in Pribram 1991, Epilogue). Carlton and I began to explore the application to brain processing of groups (Pribram & Carlton 1987), but I felt a lack of sophistication on our part regarding the psychological, perceptual process we were trying to explain. We had half the story, the how but not the what. I arranged with Roger **SHEPARD** to have him take Carlton under his wing – the result was more than I had dared hope for (see review in Pribram 1991, Lecture 5, regarding both the what and the how of object perception).

The beauty of DeValois and DeValois' experimental/theoretical program of the "how" of color perception is not only the clarification of how we see blue, but also that the same neural components of circuits can account for the perception of both color and form. Depending on the challenge to the system, either color or form, or both, can result from the process.

Martha Wilson's program of research (Wilson 1987; Wilson & DeBauche 1981) was based on Harry Helson's (Wilson's father's) adaptation level concept. Wilson showed that after removal of the inferotemporal cortex, primates failed to develop an adaptation level against which other stimuli could be compared.

Within-category discrimination may be improved by the fact that extreme values on the continuum are more identifiable and hence more discriminable when they enter into stimulus pairs. This type of salience follows from adaptation-level theory as a consequence of the privileged status of stimuli that are the most distant from the indifference point and thus appear to be prototypical stimuli as described by Rosch (1973). Adaptation-level theory thus provides a conceptual bridge between categorical perception as studied in psychophysical research and studies of prototypes and exemplars (Streitfeld & Wilson 1986, p. 449).

Wilson's further experiments and those performed on humans (Grossman & Wilson 1987) have shown that two types of categories can be distinguished: those that are image and object-form driven, such as hue and shape, and those that are comprehension-driven, such as fruit and vegetables. In monkeys, hemispheric specificity with respect to categorization has not been tested as yet; in humans, image and object-form-driven categorization is disrupted by right hemisphere lesions; left hemisphere lesions disrupt comprehension-driven categorization. Furthermore, category boundaries are disrupted by posterior, not frontal, lesions. Conversely, it is frontal lesions and not posterior ones that impair within-category performance (Pribram 1991, pp. 178–79).

The results of my program of research are consonant with those developed by Wilson. I showed that brain processes based on both convolution (using inner products of scalers) and matrix (using outer products of vectors) are involved in cognitive processing depending on whether primate frontolimbic systems or the cortex of the posterior cortex is addressed (Pribram 1986; reviewed by Pribram 1991, Lecture 10). Yasue, Jibu, and I (Pribram 1991, Appendices F and G) developed a mathematical formulation of the "how" of the formation of prototypes and of an inference process akin to the Bayesian approach used by **SHEPARD**.

The studies reviewed briefly here are attempts at providing the "how" by which **SHEPARD**'s representational process can be implemented by the brain on a day-to-day basis. Together with Shepard's program, a sophisticated set of studies has become available to coordinate psychological process, mind, with operations of brain.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

# Regularities, context, and neural coding: Are universals reflected in the experienced world?

Antonino Raffone,[a] Marta Olivetti Belardinelli,[a,b] and Cees van Leeuwen,[a,c]

[a]*Department of Psychology, University of Sunderland, St. Peter's Campus, SR6 0DD Sunderland, United Kingdom;* [b]*ECONA-Interuniversity Centre for Research on Cognitive Processing in Natural and Artificial Systems, I 00185 Rome, Italy; and* [c]*RIKEN BSI, 2-1 Hirosawa, Saitama 351–0198, Japan.*
**antonino.raffone@sunderland.ac.uk** **belarditecona@iol.it**
**cees.van-leeuwen@sunderland.ac.uk**

**Abstract:** Barlow's concept of the exploitation of environmental statistical regularities may be more plausibly related to brain mechanisms than Shepard's notion of internalisation. In our view, Barlow endorses a bottom-up approach to neural coding and processing, whereas we suggest that feedback interactions in the visual system, as well as chaotic correlation dynamics in the brain, are crucial in exploiting and *assimilating* environmental regularities. We also discuss the "conceptual tension" between Shepard's ideas of law internalisation and evolutionary adaptation.
**[BARLOW; SHEPARD]**

## BARLOW'S ARTICLE

**BARLOW**'s article well-addresses the problem of adaptation to (or internalisation of) environmental regularities and provides a relevant historical background for this problem, relating perceptual principles to neural processes. We agree with BARLOW on the relevance of the neurophysiological mechanisms involved in dealing with environmental regularities, including the perceptual experiences observed in **SHEPARD**'s elegant experiments.

**BARLOW** stresses the statistical logic of different processing stages in visual perception and object recognition, reflecting the computational exploitation of the environment's statistical regularities. In our view, the *associative*, *discriminative*, and *predictive* neural operations, which according to Barlow involve visual event occurence and co-occurence estimates, may be uniformly based on *chaotic* correlation neurodynamics. We suggest that the degree of experiential correlation between visual features or events is translated into *synchronisation grades* between chaotic neurons or neural assemblies (Raffone & van Leeuwen, in preparation; Van Leeuwen et al. 1997).

In **BARLOW**'s view, neural representations are "labelled" by the relative number and firing rates of the neurons involved. By contrast, economic representations in the cortex which exploit visual input redundancy may be given by *graded* and *intermittent* neural synchronisation patterns, enabling neurons to participate in multiple computations during the same time period. The relative degrees of correlation between neuronal discharges would effectively represent the "co-occurrence frequency" between the features or events coded by different neurons. The "independent occurrence frequency" would be coded by the number and/or firing rate of neurons coding for unrelated elementary features.

Since according to this functional logic, neurons may be involved in coding several unrelated patterns through a nontransitive short-lived synchronization (fast global inhibition may contribute to *between-pattern segregation*), anti-Hebbian synapses (Barlow 1992) or locally acting decorrelation mechanisms may not be necessary. Moreover, perceptual and memory processes may be governed by uniform neural dynamics reflecting *chaotic itinerancy*. According to Tsuda (in press), chaotic itinerancy arises when an intermediate state between order and disorder appears, and the dynamics of such a state may be regarded as those of an itinerant process, indicating a correlated transition among states. Such an itinerant process often becomes chaotic. Pattern association, discrimination, and prediction may then be expressed by relative correlation strengths in state transitions. Local *structuring* forces in terms of Hebbian coupling adaptivity and, at a more global level, *decorrelating* chaos may operate together in visual segmentation, pattern recognition, and retention (Van Leeuwen et al. 1997; Van Leeuwen & Raffone 2001).

Furthermore, **BARLOW**'s article stresses the priority of novel (unexpected) events or co-occurrences in terms of neural responses. We believe, however, that such a priority is context and time-scale dependent. Repeatedly co-occurring sensory events may be related to meaningful entities which may prevail over noisy environmental patterns. Adaptivity or sensitization of neuronal responses may be crucially dependent on the context. In contrast to Barlow's essentially bottom-up approach to neural coding, we suggest that environmental regularities are *assimilated* into pre-existing schemata rather than being simply extracted by the incoming data. Visual receptive fields may be modulated by the context (e.g., Zipser et al. 1996), as well as exhibit non-stationarity or time-dependence (e.g., Dinse 1990). These top-down properties of receptive fields are dependent on the dense patterns of feedback connectivity in the visual cortex.

As a result, neural hypotheses on the external world state may be generated "within" the cortical networks, rather than being computed in terms of mere statistical analyses of the external signals. Regularities may be *created within* the brain instead of being independently given, their effective computation may depend on the behavioural state, and they may emerge in a non-incremental or dynamically discontinuous manner. The aforementioned chaotic neurodynamic patterns may be crucial in exploiting and assimilating environmental regularities, as well as in enabling the context-sensitive processing of "spatio-temporally local" facets of the external world.

We agree with **BARLOW** on the plausibility of rotation-tuned visual cortical neurons, which may be involved in the twisting motion experience in **SHEPARD**'s experiments. Barlow points out the relevance of movement *interpolation* and *extrapolation* computations in the visual cortex, in terms of the serial involvement of different visual areas, from V1 to extrastriate areas. However, it may be that these computations are operated in terms of recursive interactions between V1 (with the highest spatial resolution) and extrastriate visual areas (with large receptive fields). Interpolation may be computed in V1 according to extrastriate (more spatio-temporally global) cues, and extrapolation may be mainly operated in extrastriate areas, given (spatio-temporally local) movement input from V1. Hence, cortical feedback may be necessary for the continuity of visual experience in space and time.

Finally, considering the *complexity* of organism-environment interactions, the Shannonian (Shannon & Weaver 1949) notion of information, on which **BARLOW**'s view is based, may not be adequate, since Shannon's information does not deal with meaning-related or contextual aspects of the interaction with the external world (Atmanspacher & Scheingraber 1991). In fact, information exploitation should depend on the brain's *understanding* (semantics) and *use* (pragmatics) of "messages," which may not be entirely specified by the actual signal structure (syntax). The notions of "signal" and "information" must not be confused (Von Foerster 1982). From a complex dynamic system viewpoint, in which the information receiver (organism) and source (environment) are not seen as dynamically separate entities, self-referent, context-dependent, and co-operative interactions dominate over separate input and output messages (actions) (Haken 1988; Olivetti Belardinelli 1976). Furthermore, it has been shown that input signals may *modulate* intrinsic correlations in the brain, rather than being in themselves meaningful (Tononi et al. 1996). In this view, the brain may go "beyond the information given."

## SHEPARD'S ARTICLE

**SHEPARD**'s experiments discussed in the target-article, eloquently demonstrate general computational principles operating in perception and cognition. However, we point out that **SHEPARD**'s general view, according to which perceptual-cognitive universals are regarded as reflections of the world throughout evolution, implies two relevant conceptual problems.

First, stating that the physical description of the outer world is



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

governed by universal rather than context-dependent or relativistic principles, may be controversial in the light of the recent theories of complex systems (e.g., Arecchi 2000; Prigogine & Nicolis 1987). The importance of achieving a synthesis between realism and relativism has also emerged in recent developments in the philosophy of science, such as in Putnam's (1980) notion of an "internal realism." "Universal" and "objective" criteria may only be defined within a given level of reality, and thus a complementary relativist assumption is implied.

Second, **SHEPARD**'s idea that evolution may have led to the internalisation of the same perceptual and cognitive principles in the mind of different subjects (organisms), and even of different planets, does not consider that adaptive interactions are plausibly based on *perception/action loops.* What may be the adaptive value of *contemplating* the universality, invariance, and elegance of the principles which govern the external world? If we assume that something beyond mere contemplation, for instance, actions related to the perceptual information, may be required in order to adapt to a spatially and temporally local environment, then the structural and dynamic properties of the sensory and effector systems must play a crucial role. But these properties may significantly differ between organisms in the universe, due to varying ecological constraints. Thus, SHEPARD's notion of evolutionary internalisation may not be a plausible general bio-cognitive principle.

**SHEPARD** has indicated general and unitary principles underlying perception and cognition. We do believe that searching for such principles may be extremely significant for cognitive psychology, where often only "local" or "domain-specific" accounts are put forth. For instance, Shepard's so-called "universal law of generalization" may be reflected in general quantitative principles of perceptual grouping or Gestalt principles, in terms of an "attraction law" (Kubovy & Wagemans 1995). However, rather than assuming that the mind performs mathematical calculations in terms of spatial coordinates or static representations, we believe that perceptual (cognitive) laws are reflected in subsymbolic dynamics and emergent neural patterns. According to this view, evolution and learning may operate in a synergic manner, as stressed in **TENENBAUM & GRIFFITHS'** article, just as proximity and previous knowledge may interact in visual and auditory organisation.

## An alternate route toward a science of mind

David A. Schwartz

*Department of Psychology, University of North Carolina-Asheville, Asheville, NC 28804.* **dschwartz@unca.edu**

**Abstract:** Shepard has challenged psychologists to identify nonarbitrary principles of mind upon which to build a more explanatory and general cognitive science. I suggest that such nonarbitrary principles may fruitfully be sought not only in the laws of physics and mathematics, but also in the logical entailments of different categories of representation. In the example offered here, conceptualizing mental events as indexical with respect to the events they represent enables one to account parsimoniously for a wide range of empirical psychological phenomena.
**[SHEPARD]**

Each of the foregoing target articles addresses one or more of **SHEPARD**'s specific theoretical hypotheses, such as that the phenomenon of apparent motion reflects the internalization of abstract principles of kinematic geometry, or that empirical generalization gradients reflect the internalization of principles of Bayesian inference. SHEPARD's internalization hypothesis may or may not turn out ultimately to be the most useful interpretation of these phenomena, but his example of a sustained attempt to explain, rather than simply to describe, the observed regularities of cognitive functioning surely deserves wider emulation. In this commentary I offer a very different example of how the search for nonarbitrary principles governing cognitive functioning can support the theoretical integration of widely disparate empirical phenomena.

Cognitive science can fairly be defined as the study of mental representation, and debates within the literature over the nature of representational processes are often framed as contests between members of a pair of binary opposites, for example, analog vs. digital, propositional vs. imagistic, modal vs. amodal. Students of semiotics, however, recognize three qualitatively different categories of representation, which, following Peirce (1966), they call *symbol, icon,* and *index,* respectively. Within the technical discourse of semiotics, the term *symbol* refers to a sign that bears a purely arbitrary and conventional relationship to the object it represents. All non-onomatopoetic words in a natural human language are symbols. The term *icon* refers to a sign that bears some physical resemblance to the object it represents. A pictograph is an icon of its referent, as is a portrait of the individual who sat for it. Finally, the term *index* refers to a sign that bears a physical causal relationship to the object it represents, such that the object is the cause and the index is the effect. Examples of indices include weathervanes (which indicate wind direction), smoke (which indicates fire), and spontaneous facial expressions (which indicate an individual's internal mood state).

While the experienced content of much human thought is arguably iconic (Barsalou 1999), and the signs humans use to communicate their thoughts to each other are often symbolic, the proposition that mental events are indexical with respect to the events they signify turns out to be an especially fertile point of departure in the search for nonarbitrary principles governing cognitive functioning. Consider that the cause-and-effect relationship between an index and the object it signifies entails the following three properties: (1) an index necessarily signifies presence; (2) an index cannot explicitly signify absence; and (3) an index is veridical; it cannot lie (though it can be misinterpreted). By contrast, the arbitrariness of symbolic representations makes it possible to communicate about the absent, the hypothetical, the counterfactual, as well to transmit false information (Rappaport 1999).

The premise that mental events are indexical with respect to the events they represent thus implies that: (1) mental events necessarily signify some positive state of affairs; (2) they cannot explicitly represent absence or nonexistence; and (3) they are necessarily veridical. That at least some mental events exhibit these indexical properties is clearly evident in the phenomenon of perception. By one definition, perception is "the consciousness of particular material things *present* to sense" (James 1892, p. 179; emphasis added). A corollary to this definition is that one cannot perceive absence, although one can note a mismatch between what is perceived and what was expected. Implication (3) leads directly into thorny philosophical terrain, for the veridicality of perception, or of any mental event, is notoriously difficult to establish. As **SHEPARD** himself has reminded us, we cannot step outside our minds in order to evaluate the correspondence between our thoughts and "reality" (Shepard 1981b). The most we can safely claim, therefore, is that, whatever might be the actual veridicality of perceptual events, we seem to have evolved to behave as if our perceptions were indeed valid indicators of the state of the world and of our own bodies.

Those authors who view perception and cognition as qualitatively different psychological processes might feel inclined to argue that the apparent indexicality of perceptual events has no bearing on the attempt to explain regularities of cognitive functioning. This argument appeals especially to those who view cognition as fundamentally propositional in nature, involving the manipulation of arbitrary symbols. For this very reason, findings that patently cognitive phenomena do indeed exhibit the indexical properties characteristic of perception, should call into question any alleged strong separation between perceptual and cognitive processes. The hypothesis that cognitive phenomena exhibit indexical properties can be reformulated into a number of more specific empirical predictions including: (1) individuals will exhibit


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

greater sensitivity to present than to absent stimuli; (2) individuals will exhibit relative difficulty comprehending or employing truth-functional negation (because representing negation entails representing the absence of that which is negated); and (3) individuals will tend to treat mentally represented propositions as true, inasmuch as individuals have evolved to behave as if the mental presence of a thought indicates the reality of the state of affairs to which the thought refers.

There exists a considerable body of empirical evidence consistent with the above predictions. I include here only a representative sample of the relevant findings. For a more extensive review see Schwartz (1998). Regarding individuals' relative insensitivity to absence, psychophysicists have found that people detect the appearance (onset) of a visual (Bartlett et al. 1968), auditory (Zera & Green 1993), or tactile (Sticht & Gibson 1967) stimulus more readily than they detect the disappearance (offset) of the stimulus. Similarly, many animal species learn contingencies involving presence more readily than they learn those involving absence. For example, pigeons and nonhuman primates discriminate much better between two stimuli in feature-positive situations (in which the presence of something signals a reward) than in feature-negative situations (in which the absence of a feature signals a reward; Hearst 1991). Finally, social psychologists have documented that people have more difficulty detecting covariations when the presence of one stimulus covaries with the absence of another than when it covaries with the presence of another (Nisbett & Ross 1980).

Regarding individuals' relative difficulty representing truth-functional negation, Wason (1959; 1966), reported that people find it easier to reason with affirmative statements (e.g., modus ponens) than with denials (e.g., modus tollens). Clark (1974), moreover, showed that people find negation more difficult to comprehend than affirmation, and that in sentence verification tasks, people take longer to verify denials ("*x* is false") than to verify affirmatives ("*x* is true"). Finally the ability to deny propositions (i.e., to employ truth-functional negation) is one of the last linguistic abilities to emerge in childhood, suggesting that representing negation poses particular challenges to a developing mind (Gilbert 1991; Horn 1989).

Regarding the prediction that individuals will tend to treat mentally represented propositions as true, psychologists have shown: (1) that belief is automatic upon comprehension (Gilbert 1991); (2) that beliefs often persist even when explicitly discredited ( Johnson & Seifert 1994; Nisbett & Ross 1980); and (3) that the strength of a belief often varies as a function of one's familiarity with the information a given proposition contains (Arkes 1991). These phenomena are consistent with an indexical hypothesis whereby: (1) experience leaves physical traces in the mind; (2) these traces, once formed, are not easily removed; and (3) the stronger the trace, the stronger the feeling of reality or validity associated with it.

This effort to explain certain regularities of cognitive functioning as logical entailments of indexicality offers none of the formal mathematical elegance of the work of **SHEPARD** and his colleagues, in part because it proceeds from a conceptual framework (i.e., semiotics) that is itself not well formalized. It does, however, demonstrate that the attempt to identify nonarbitrary principles of mind can extend well beyond the phenomena of perceptual organization or categorization, and offers a very different example of how one might pursue the goal of a more explanatory and general cognitive science.

## Sphericity in cognition

E. N. Sokolov

*Department of Psychophysiology, Faculty of Psychology, Moscow Lomonosov State University, Moscow, 103009, Russia.* **ensok@mail.ru**

**Abstract:** The perceptual circularity demonstrated by R. Shepard with respect to hue turns out to be a sphericity of color perception based on color excitation vectors of neuronal level. The spherical color model implicitly contains information concerning generalization under color learning. Subjective color differences are "computed" in neuronal nets being represented by amplitudes of evoked potentials triggered by color change. **[SHEPARD]**

**R. SHEPARD** in a very seminal paper emphasizes that there are circular components in perception, a feature that has been historically ignored by most of psychophysics. Illuminating the role of circularity, **SHEPARD** refers to color space as a three-dimensional spherical solid. Boring (1951) has suggested a four-dimensional color solid. Following this line of research, I have found that color space is not a solid, but a surface – a hypersphere in the four-dimensional Euclidean space (Izmailov & Sokolov 1991). Cartesian coordinates of points representing colors on the hypersphere correspond to excitations of four types of neurons present in the primate lateral geniculate body: red-green, blue-yellow, brightness, and darkness cells. Thus, colors are encoded by excitation vectors of a constant length. The spherical coordinates (three angles of the hypersphere) correspond to subjective aspects of color perception: hue, lightness, and saturation. It means that the circular representation of hue emphasized by **SHEPARD** is only one circular phenomenon of color perception. Lightness and saturation are also circular, as are other angles of the hypersphere.

Specific colors on the hypersphere are represented by color detectors situated in the area V4 of the primate visual cortex. Subjective differences between colors correspond to distances between respective color detectors. These distances are measured, however, not along geodesic lines, but as Euclidean distances between the ends of excitation vectors encoding respective colors. This implies that subjective differences highly correlate with absolute values of respective vectoral differences (Izmailov & Sokolov 1991). It suggests that color differences are "computed" in specific networks. The evidence for this was found in humans by recordings of cortical evoked potentials elicited by color change. The amplitudes of the N87 component elicited by substitution of color stimuli correlate highly with subjective differences between respective colors (Izmailov & Sokolov 2000).

A matrix of N87 amplitudes analyzed by multidimensional scaling (MDS) revealed a hypersphere in the four-dimensional space coinciding with the four-dimensional space found by MDS of the matrix for subjective color differences (Izmailov & Sokolov 2000).

The color differences "computed" in neuronal networks explain the hyperbolic function of reaction time (RT) by detection of color targets. Comparison of RTs and subjective color differences demonstrates their reciprocal relation. The greater the color difference, the shorter is RT – approaching a non-reducible minimum (Sokolov 1998).

The correspondence between generalization gradients and Euclidean distances found by **SHEPARD** for hue can be extended to the four-dimensional color space. Matrices of response probabilities obtained under differential color conditioning revealed a hypersphere in the four-dimensional space for rhesus monkey and carp. This relationship is due to the fact that color conditioning results from plastic modifications of synapses on a command neuron under the influence of the color excitation vector of the conditional stimulus. A set of plastic synapses on a command neuron constitutes a link vector. In the process of learning, the synaptic link vector becomes equal to the excitation vector of the conditional stimulus making a command neuron selectively tuned to the conditional stimulus. Command neurons summating products of multiplications of synaptic weights by respective component of the excitation vectors, "compute" an inner product of the established



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

link vector and excitation vectors of differential color stimuli. The greater the deviance of the differential stimulus from the conditional one, the smaller the inner product and the lower the response probability (Sokolov 2000). This results in a reciprocal relationship between color difference and response probability, in accordance with Shepard's generalization theory.

A very important argument for the universal character of the four-dimensional spherical model of color space was obtained by intracellular recordings from bipolar cells of carp retina. Six types of tonic bipolar cells were found: four types of opponent cells (red+green; green+red; yellow+blue; blue+yellow) and two types of non-opponent cells (brightness and darkness). Opponent color cells work on an "either-or" basis, so that only two cells can be simultaneously activated. Two non-opponent cells are activated to a certain degree, instantly. Thus, a maximum of four types of color-coding cells can be active at once. For all wavelengths of colors the sum of squared amplitudes of the four types of cells was equal to a constant value, demonstrating the sphericity of four-dimensional color space (Chernorizov & Sokolov 2001).

## Shepard's mirrors or Simon's scissors?

Peter M. Todd and Gerd Gigerenzer

*Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, 14195 Berlin, Germany.* **ptodd@mpib-berlin.mpg.de**
**www-abc.mpib-berlin.mpg.de/users/ptodd**
**gigerenzer@mpib-berlin.mpg.de**
**http://www.mpib-berlin.mpg.de/ABC/Staff/gigerenzer/home-d.htm**

**Abstract:** Shepard promotes the important view that evolution constructs cognitive mechanisms that work with internalized aspects of the structure of their environment. But what can this internalization mean? We contrast three views: Shepard's mirrors reflecting the world, Brunswik's lens inferring the world, and Simon's scissors exploiting the world. We argue that Simon's scissors metaphor is more appropriate for higher-order cognitive mechanisms and ask how far it can also be applied to perceptual tasks.
**[BARLOW; KUBOVY & EPSTEIN; SHEPARD]**

***What's in the black box?*** To understand the contents of the mind, we should consider the environment in which it acts and in which it has evolved. **SHEPARD**'s work has done much to spread this important ecological perspective, focusing on a particular vision of how the external world shapes our mental mechanisms. For SHEPARD, much of perception and cognition is done with mirrors: key aspects of the environment are internalized in the brain "by natural selection specifically to provide a veridical representation of significant objects and events in the external world" (SHEPARD, this issue, p. 582). Without entering into arguments over the need for representations of any sort (see e.g., Brooks 1991a), we can still question whether representations should be veridical, constructed to reflect the world accurately, or, instead, be useful in an adaptive sense. Clearly, not all veridical representations are useful, and not all useful representations are veridical. A less exacting view of internalization can be seen in the work of Egon Brunswik (as discussed by **BARLOW**, this issue), who proposed a "lens model" that reconstructs a representation of a distal stimulus on the basis of the uncertain proximal cues (whose availability could vary from one situation to the next) along with stored knowledge of the environmental relationships (e.g., correlations) between those perceived cues and the stimulus (Brunswik 1955). For Brunswik, the mind infers the world more than it reflects it. Herbert Simon expressed a still looser coupling between mind and world: bounded rationality, he said, is shaped by a pair of scissors whose two blades are the characteristics of the task environment and the computational capabilities of the decision maker (Simon 1990). Here, the mind must fit closely to the environment, but the two are complementary, rather than mirror images.

We expect that the mind draws on mechanisms akin to all three tools, mirrors, lenses, and scissors, from its adaptive toolbox (Gigerenzer & Todd 1999a). The question now becomes, where can each be applied? In perception, using Shepard's mirror or Brunswik's lens may often be the right way to look at things, but there are also instances where these tools are inappropriate. Consider the problem of a fielder trying to catch a ball coming down in front of her. The final destination of the ball will be complexly determined by its initial velocity, its spin, the effects of wind all along its path, and other causal factors. But rather than needing to perceive all these characteristics, reflect or model the world, and compute an interception point to aim at (with screw displacements or anything else), the fielder can use a simple heuristic: fixate on the ball and adjust her speed while running toward it so that her angle of gaze – the angle between the ball and the ground from her eye – remains constant (McLeod & Dienes 1996). By employing this simple gaze heuristic, the fielder will catch the ball while running. No veridical representations or even uncertain estimates of the many causal variables in the world are needed – just a mechanism that fits with and exploits the relevant structure of the environment, namely, the single cue of gaze angle. How widely such scissors-like heuristics can be found in perception remains to be seen, but some researchers (e.g., Ramachandran 1990) expect that perception is a "bag of tricks" rather than a box of mirrors.

***Extending an ecological perspective to higher-order cognition.*** When we come to higher-order cognition, Simon's cutting perspective seems the most appropriate way to extend **SHEPARD**'s ecological view. Consider a simple cognitive strategy that has been proposed as a model of human choice: the Take The Best heuristic (Gigerenzer & Goldstein 1996). To choose between two options on the basis of several cues known about each option, this heuristic says to consider one cue at a time in order of their ecological validity, and to stop this cue search with the first one that distinguishes between the options. This "fast and frugal" heuristic makes decisions approximately as well as multiple regression does in many environments (Czerlinski et al. 1999), but usually considers far less information (cues) in reaching a decision. It does not incorporate enough knowledge to reasonably be said to reflect the environment, nor even to "model" it in Brunswik's sense (because it knows only cue order, not even exact validities), but it can certainly match and exploit environment structure: When cue importance is distributed in an exponentially decreasing manner (as often seems to be the case), Take The Best cannot be outperformed by multiple regression or any other linear decision rule (Martignon & Hoffrage 1999). In this situation, the two scissor blades cut most effectively. As another example, the QuickEst heuristic for estimating quantities (Hertwig et al. 1999) is similarly designed to use only those cues necessary to reach a reasonable inference. QuickEst makes accurate estimates with a minimum of information when the objects in the environment follow a J-shaped (power law) distribution, such as the sizes of cities or the number of publications per psychologist. Again this crucial aspect of environment structure is nowhere "built into" the cognitive mechanism, but by processing the most important cues in an appropriate order, QuickEst can exploit that structure to great advantage. Neither of these heuristics embodies logical rationality – they do not even consider all the available information – but both demonstrate ecological rationality, that is, how to make adaptive decisions by relying on the structure of the environment.

Why might Simon's scissors help us understand cognitive mechanisms better than **SHEPARD**'s mirror? We (and others) suspect that humans often use simple cognitive mechanisms that are built upon (and receive their inputs from) much more complex lower-level perceptual mechanisms (Gigerenzer & Todd 1999a). If these heuristics achieve their simplicity in part by minimizing the amount of information they use, then they are less likely to reflect the external world and more likely to exploit just the important, useful aspects of it, as calculated and distilled by the perceptual system (which may well base its computations on a more reflective representation).[1] While **KUBOVY & EPSTEIN** (this issue) would probably argue that neither metaphor, mirrors or scissors, helps us in specifying cognitive mechanisms, we feel that such metaphors


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

are vital in guiding research by providing an image of the sort of mechanisms to seek (as has been the case throughout the history of psychology – see Gigerenzer 1991). This is why it is important to point out that Simon's scissors may be a better model to have in mind than Shepard's mirror when studying a range of mental mechanisms, particularly higher-level ones.

Thus, in extending **SHEPARD**'s search for the imprint of the world on the mind from perception to higher-order cognition, we should probably look less for reflections and more for gleams. To achieve this extension, we must also discover and consider more of the "general properties that characterize the environments in which organisms with advanced visual and locomotor capabilities are likely to survive and reproduce" (SHEPARD, this issue, p. 581); these might include power laws governing scale invariance (Bak 1997), or principles of adaptively unpredictable "protean behavior" (Driver & Humphries 1988), or dynamics of signaling between agents with conflicting interests (Zahavi & Zahavi 1997), or costs of time and energy in seeking information (Todd 2001). With characteristic structures such as these before us as one half of Simon's scissors, we can look more effectively for the cognitive mechanisms that form the other, matching half.

NOTE

**1.** This is not to say that simplicity and frugality do not also exert selective pressure on perceptual mechanisms – **SHEPARD** appreciates the need for simplicity and speed of computation in those systems as well, for instance proposing screw displacement motions as representations because they are "geometrically simplest and hence, perhaps, the most quickly and easily computed" (Shepard, this issue, p. 585). But the amount and manner of information and processing may differ qualitatively from that in higher-order cognitive mechanisms.

## Measurement theory is a poor model of the relation of kinematic geometry and perception of motion

Dejan Todorovič

*Department of Psychology, University of Belgrade, 11000 Belgrade, Serbia, Yugoslavia.* **dtodorov@dekart.f.bg.ac.yu**

**Abstract:** The Kubovy-Epstein proposal for the formalization of the relation between kinematic geometry and perception of motion has formal problems in itself. Motion phenomena are inadequately captured by the relational structures and the notion of isomorphism taken over from measurement theory.
**[KUBOVY & EPSTEIN]**

**KUBOVY & EPSTEIN (K&E)** use measurement theory as a model to couch the relation of kinematic geometry and perception of motion in more formalist terms than **SHEPARD**. A virtue of successful formalization is the conceptual organization and clarification of a group of phenomena through succinct expression of their essential aspects. However, to be appropriate, the tools of formalization must adequately mirror the corresponding empirical domain. As I will argue, measurement theory falls short as a model for motion phenomena.

Measurement may be formalized by singling out some physical entities and procedures involved in the empirical process of taking measures, and identifying their mathematical counterparts. **K&E** use the example of formalizing weight measurement. In this case, on the physical side one would consider, first, some objects having weights, second, a physical procedure of comparing the weights of two objects, and third, another physical procedure of putting two objects together so their weights combine. In counterpart, on the mathematical side, there are, first, numbers corresponding to the numerical values of weights, second, the mathematical relation of comparison of two numbers, and third, the mathematical operation of addition of two numbers. When the physical objects are considered together with the above procedures, they form a particular logical "relational structure"; similarly, the mathematical objects and procedures form another relational structure. I will call "measurement domains" such logical structures whose constituents are a set of elements, a relation of comparison, and an operation of composition. The motivation for considering such entities is the construction of a mathematical measurement domain as a model of a physical measurement domain. This is accomplished when, as in the above example, the two domains are homomorphic, that is, when their corresponding constituents map one-to-one onto each other, so that their logical structures are equivalent.

**K&E** consider two physical and two perceptual domains. The domain of "kinematic geometry" is the mathematical counterpart of the domain of "physical motions," whereas the domain of "models of perception" is the counterpart of the domain of "perception of motion." **K&E** claim, first, that the four domains are measurement domains in the above sense, and second, that the relations between them are homomorphisms (see their Fig. 2). I dispute both claims and argue that the proposed formalization is inadequate.

Consider first the structures of the domains. As the structure of the perceptual domains is somewhat unclear, I will concentrate on the physical domains. To establish that a domain of interest is a measurement domain, what is needed is to identify its constituents, that is, the elements and the appropriate relation and operation. In the domain of "physical motions," the elements are presumably objects that can exhibit motions. But what could be the appropriate relation of comparison of two motions, analogous to the comparison of weights of objects A and B? The problem is that "motion" is a far richer notion than "weight": In comparing motions of objects A and B, should one consider their speed, path length, duration of motion, shape of trajectory, or manner of motion along the trajectory? How do two motions compare if object A takes a shorter time but traverses a longer extent than object B, or if object A accelerates from a slower speed whereas object B decelerates from a faster speed, or if object A moves on a rectilinear path whereas object B moves on a circular path? One might choose some particular criterion for particular comparison purposes, but the problem is to define a reasonable, general way to decide which of two arbitrary motions is "greater than" the other. However, in contrast to weight comparisons, where such a decision is always possible, forcing motion comparisons into the template of a "greater than" type of relation does not appear to be generally useful. A similar problem would apply for the definition of the appropriate operation of motion composition. Since a measurement domain must exhibit an appropriate relation and operation, these formal problems suggest that "physical motions" is not one.

Related problems arise in the analysis of the logical structure of the domain of "kinematic geometry." In the weight measurement model, the elements in the mathematical relational structure are single numbers, the numerical values of weights. However, scalars are inadequate to express motions. Even for the motion of a single point in space one needs a temporally varying 3-D vector, that is, three infinite sets of numbers. This again shows that measurement domains are poor models for describing motions.

One way to avoid these difficulties is to concede that the particular relational structure taken over from the measurement model is indeed inadequate to analyze motions, but to claim that this problem can be amended by constructing a more elaborate and adequate logical structure. Such a structure should be appropriately suited for motions, and its constituents would involve the notions of shape and length of trajectories, speed, uniform and non-uniform motion, acceleration, and so on. However, such a new structure would not, in my judgment, be substantially any different from the already existing physical theories of motion, that is, kinematics and dynamics, and thus would provide little conceptual advance.

Consider now the structural relation between the domains of "physical motions" and "perception of motion," which, according



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

to K&E, is a homomorphism. When a homomorphism exists between two domains, this means that they are essentially formal mirror images of each other. But are physical motions and our perceptions of them indeed so closely related? This could be the case if perception of motion were veridical. However, it is well known that this is not generally true: motions of equal lengths can be perceived to have different lengths (Mack & Herman 1972); motions of equal speeds can be perceived to have different speeds (Loomis & Nakayama 1973); rectilinear motions can be perceived to be curvilinear (Johansson 1950); objects at rest can be perceived as moving (Duncker 1929). Such data indicate that the structural relation between physical motions and motion perception is not likely to be a homomorphism; consequently, the claim by K&E that kinematic geometry is a model for perception of motion is dubious. A way to circumvent these problems could be to concede that homomorphism does not apply in this case, and to try to formulate a new between-domain structural relation instead; one more complicated than homomorphism, that would more adequately describe the relation between physical motion and motion perception. However, if such a description were to amount to a restatement of the already known empirical findings of motion research – and it is hard to see what else it could do – then it would serve little purpose.

In sum, neither the notion of relational structure nor the notion of homomorphism, as used in measurement theory, do formal justice to the studies of physical and perceived motion. It still remains to be shown that a formalization of their relation is feasible and useful.

## Minimization of modal contours: An instance of an evolutionary internalized geometric regularity?

Giorgio Vallortigara[a] and Luca Tommasi[b]

[a]*Department of Psychology and B.R.A.I.N. Centre for Neuroscience, University of Trieste, 34123 Trieste, Italy;* [b]*Department of General Psychology, University of Padua, 35131 Padua, Italy.*

**vallorti@univ.trieste.it** **http://www.psico.univ.trieste.it/labs/acn-lab/**
**ltommasi@psico.unipd.it** **http://www.psy.unipd.it/~ltommasi/**

**Abstract:** The stratification in depth of chromatically homogeneous overlapping figures depends on a minimization rule which assigns the status of being "in front" to the figure that requires the formation of shorter modal contours. This rule has been proven valid also in birds, whose visual neuroanatomy is radically different from that of other mammals, thus suggesting an example of evolutionary convergence toward a perceptual universal.

**[SHEPARD]**

We will contribute briefly to the hypothesis that the brain exploits (and embodies) regularities in the environment with one of the few examples, we believe, in which a specific geometric regularity has been put forward as the explanation of certain perceptual phenomena in humans, and also tested for its generality across species. A basic computational problem in perceiving occlusion deals with establishing the direction of depth stratification, that is, which surface is in front and which is behind. Usually, when two objects differ in colour, brightness, or texture, occlusion indeterminacy can be solved using T-junctions (see Cavanagh 1987), that is, by determining, on the basis of contour colinearity, what boundaries belong to each other and thereby allowing the formation of modal (occluding) and amodal (occluded) contours (see Kanizsa 1979; Michotte 1963; Michotte et al. 1964). However, humans can perceive unconnected and depth-stratified surfaces even in chromatically homogenous patterns, with only L-junctions and no T-junctions at all. Although it would be possible, in principle, to perceive a unitary object, in the figure provided (see Forkman & Vallortigara 1999) the hen appears to be behind the fence when the region of the legs is inspected (because of the differences in colour that specify the direction of occlusion), but it appears to be in front of the fence when the region of the trunk is inspected (Fig. 1).

Why is it that the trunk of the hen appears for most observers (and/or for most of the time) in front of the fence rather than the other way round? The reason why larger surfaces (such as the hen's trunk) tend to be seen modally in front, rather than behind, is because of the geometrical property that when in overlapping objects larger surfaces are closer, there will be shorter occluding boundaries than when smaller surfaces are closer. Shorter modal (occluding) contours are needed to account for the occlusive effect of the hen on the fence, whereas larger modal contours would be needed to account for the occlusive effect of the fence on the hen's trunk. This "rule," first described by Petter (1956), according to which the visual system tends to minimize the formation of interpolated modal contours, has been largely confirmed in studies of human visual perception (Rock 1993; Shipley & Kellman 1992; Singh et al. 1999; Thornber & Williams 1996; Tommasi et al. 1995). It has been shown that the rule is independent of the empirical depth cue of relative size (Tommasi et al. 1995) and can be made to play against information based on other depth cues, thus generating intriguing visual paradoxes such as the hen/fence illusion (see Kanizsa 1979 for further examples). It is noteworthy that the interpretation we propose here is exactly that minimization of modal contours occurs to exploit a geometric regularity. An alternative view would be to suppose that it is an instantiation of a more general "minimum principle" (Cutting & Proffitt 1982). We think that the principle is indeed general, in that it would allow an organism a simple rule to approach the closest object (Zanforlin 1976), but we believe that the minimization of modal contours occurs as a way to implement the geometric regularity rather than as the accidental by-product of a general-purpose minimization process.

Is Petter's rule a geometric regularity which has been incorporated in the design of all vertebrate brains or is it limited to the human visual machinery? The use of the hen in the figure is not incidental. Forkman and Vallortigara (1999) have recently reported that domestic hens behave in visual discrimination tests as if they would experience the same phenomena. This is remarkable, we think, for although in vertebrates the general pattern of organization of the nervous system is quite conservative, the brain of birds lacks the layered organization characteristic of the mammalian isocortex. Actually, birds seem to possess neural circuits functionally comparable to those of the mammalian isocortex, but with a nuclear rather than a laminar organization (Karten & Shimizu

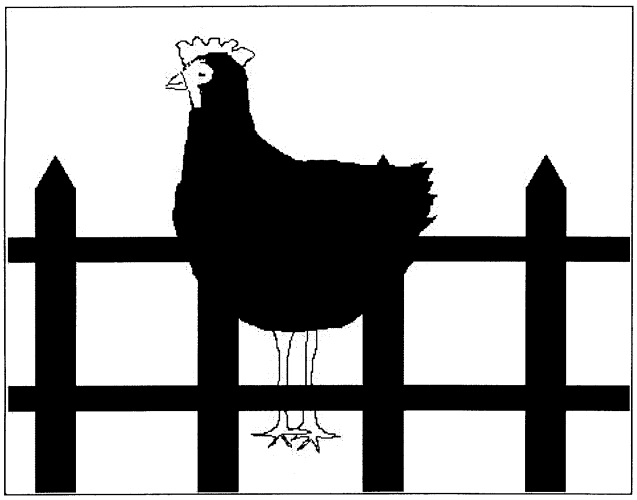


Figure 1 (Vallortigara & Tommasi). The hen appears to be standing in front of the fence in the region of the trunk, but if one inspects the region of the legs, it appears to be standing behind the fence.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

1989). As to the visual system, the pattern of connectivity and chemoarchitecture of the two major ascending visual pathways to the forebrain is closely comparable in birds and mammals (Güntürkün 1996). At present, it is unclear whether all these similarities reflect homoplasy (convergent evolution) or homology (shared evolutionary ancestry). But, whatever the answer, the possibility that some very general computational rules are conserved, or reinvented, in classes whose lineages have been separated some 250–300 million years ago is intriguing. It suggests that there are important constraints related to the geometrical and physical properties of the world that must have been incorporated in the design of any efficient biological visual system.

## Toward a generative transformational approach to visual perception

Douglas Vickers

*Psychology Department, Adelaide University, Adelaide, SA 5108, Australia.*
**Douglas.Vickers@adelaide.edu.au**
**http://www.psychology.adelaide.edu.au/members/**

**Abstract:** Shepard's notion of "internalisation" is better interpreted as a simile than a metaphor. A fractal encoding model of visual perception is sketched, in which image elements are transformed in such a way as to maximise symmetry with the current input. This view, in which the transforming system embodies what has been internalised, resolves some problems raised by the metaphoric interpretation.
**[HECHT; SHEPARD]**

**SHEPARD** has argued that the human brain has evolved in such a way as to internalise the most general principles that operate in the physical world. Although acknowledging the sweep and suggestiveness of this idea, most commentators, such as **HECHT**, question the usefulness of the metaphor of internalisation and argue that it is falsified when made specific, and unhelpful when it remains general. It is with some trepidation, therefore, that I suggest that a quite concrete instantiation of these most general principles can provide useful insights into the way in which perceptual and cognitive processes might conceivably operate.

Independently of the validity of internalisation, it seems obvious that a comprehensive and coherent explanation of the interactions between brain processes and the physical world (and between different sets of brain processes) must be in terms of a common conceptual framework. In agreement with proponents of nonlinear dynamical analyses (e.g., Port & Van Gelder 1995), I would argue that the most general common framework is that of geometry.

Within this framework, two of the most general and powerful notions for explaining and guiding our understanding of the physical universe are: (1) symmetry under some set of transformations; and (2) some form of minimum (i.e., optimisation) principle. These are the two constraining principles, in particular, that **SHEPARD** argues have been internalised. In proposing that, in default conditions, human perception tends to conform to the principles of kinematic geometry, SHEPARD attempts to map these concepts directly onto mental functioning.

I wish to suggest that a more useful interpretation of **SHEPARD**'s position is that perceptual and cognitive processes operate as if such general principles as symmetry and optimisation had somehow been internalised. In order to justify this belief, I should like to present what is still more a thought experiment than a well worked-out hypothesis. This idea takes its inspiration from recent developments in fractal geometry and, in particular, from the use of fractal encoding to compress and generate visual images. First, a very brief explanation of these ideas may be helpful.

Fractal geometry is concerned with the analysis of complex lines, surfaces, and objects that are comparable in complexity to the outlines and composition of natural objects and textures. As an example, a fractal curve can be generated by taking a simple seed element, such as a straight line (—), and making two reduced copies of it (– –). These reduced copies are subjected to a set of transformations (e.g., $\vee$). Two reduced copies are then made of each line element in this output and the same set of transformations is again applied. The process can continue indefinitely, but, in practice, five or six iterations are sufficient to approach the resolution of the system used to display (or to view) the resulting curve. More complex sets of networked probabilistic generation processes can also be used to create highly complex images.

Fractal curves and surfaces are interesting for several reasons. First, such curves and surfaces frequently resemble natural objects and environments (Mandelbrot 1983; Prusinkiewicz & Lindenmayer 1990). Second, like certain ferns, fractal objects and images exhibit "self-similarity" (i.e., their structure is statistically similar at different scales of magnification, so that a small part resembles the whole structure). Third, all the information necessary to generate a fractal curve can be represented by the parameters of a set (or collage) of some half dozen transformations.

It can be proved that any image, however complex, can in principle be represented by the parameters of a set of fractal generation processes. Because of this, as Peitgen et al. (1992, p. 259) remark, "Fractal geometry offers a totally new and powerful modelling framework for such encoding problems. In fact, we could speculate that our brain used fractal-like encoding schemes." My proposed thought experiment is that we try to work out some of the consequences of taking this suggestion seriously.

For example, as a first step in this direction, K. Preiss and I have devised a program that takes any regular fractal curve (e.g., the well-known Koch curve) and uses a plausibly constrained sequence of transformation operations to discover (by mapping the transformed patterns onto the original) the reduction factor, number of copies, translation, rotation and reflection parameters, and the number of iterations required to generate a copy of the curve. This copy may then be matched with the original. I propose that such a program can represent a highly simplified, toy model of one way that the perceptual process might conceivably operate.

To make this model more general, we might go on to speculate that the visual system carries out a collage of the simplest transformations on elements of the visual image that will maximise symmetries between the transformed output and the current image (and with changes in that image), given the rate of change in the input and the physical parameters controlling the transforming systems. That is, visual perception is conceived as the attractor-like output of a dynamic, generative transformational system that "resonates" with the current input.

From this perspective, the transforming system embodies what has been internalised. The principles of symmetry maximisation and optimisation are obvious candidates as hypotheses for describing the operation of the system. However, hypotheses in these terms are now empirically testable. Even if true as descriptions of the system, they are not used to describe the end result of perception, but the operation of the perceptual system *as it interacts with the environment.* As a result, the various examples cited by commentators as falsifying the internalisation notion, become instead empirical findings that may or may not be consistent with the predicted output of the system.

If we think of very general constraints as (possibly) applying to the system that processes external information, then the opposition between well-resolved geometric regularities and "more abstract" statistical regularities, to which **HECHT** draws attention, need not arise. Perception is now seen as determined by the distribution of active transformations that generate an output that is maximally symmetric (statistically) with the current input. That is, perceptual responses are not bound by group-theoretic requirements of perfect symmetry. At the same time, the repertoire of transformations that the system calls on may embody evolutionary developments reflecting bodily and external constraints. Thus, the system can incorporate the kind of prior knowledge (about any and every stimulus) that the Bayesian approach to perception im-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

plausibly ascribes to cognition. (It might even incorporate some aspects of HECHT's "externalisation" hypothesis.)

A further consequence of this approach is that it may be able to accommodate one of the features of responses to nongeneric views of objects that otherwise seem puzzling. This is the sudden sharp change in perception that occurs (as when a pencil is viewed end-on). To do this, however, it may be important not to underestimate the richness of the information that may be used to select the appropriate transformations.

Specifically, in our program, we used a Hausdorff distance metric to compare the transformed and the original images (Rucklidge 1996). The process of measuring the Hausdorff distance between two sets of points, A and B, takes into account the distances from each point in set A to every point in set B and the distances from each point in B to every point in A. This measuring process yields a rich landscape of information. In essence, each point in an array has a "view" of every other point.

For example, when applied to a (single) Glass figure, among the distribution of inter-point distances, one single value predominates; all that is necessary is to select the transformation that will move a point through that distance. In other words, from this perspective, some perceptual phenomena that have been seen as problematic because of a correspondence problem (which points go with which), appear to be *over-* rather than under-determined.

The generative transformational approach that I suggest is worth exploring is congruent with some other recent work in cognition, such as Feldman's (1997) treatment of one-shot categorization and Leyton's (1992) speculation that we may be sensitive to the "process-history" of objects, where this history is interpreted as a sequence of transformations from some maximally symmetric original state. However, its main merit, in the present context, is that it can provide a computational model, in terms of which **SHEPARD**'s internalisation hypothesis retains its generality. Perhaps a further appealing feature is that this is achieved by a concrete instantiation of two notions central to Shepard's thinking: transformation and a generative system.

ACKNOWLEDGMENT

I wish to thank Michael Lee for helpful comments.

## What's in a structure?

Virgil Whitmyer

*Department of Psychology, Indiana University, Bloomington, IN 47405.*
**Vwhitmye@indiana.edu**

**Abstract:** Shepard's general approach provides little specific information about the implementation of laws in brains. Theories that turn on an isomorphism between some domain and the brain, of which Shepard's is one, do not provide specific detail about the implementation of the structures they propose. But such detail is a necessary part in an explanation of mind. **[SHEPARD]**

I have a general worry about projects like **SHEPARD**'s. Consider a cousin of SHEPARD's theory which has, I take it, a similar goal. The Gestalt psychologists were out to uncover the structure of the brain. Köhler expresses a desire to explore the *"terra incognita"* between stimulation and responsive behavior (1929, p. 54). One might consider such neural territory *qua* realizer of phenomenal states, looking, for example, for opponent process cells that realize sensations of red and green, yellow and blue. On the other hand, she might be interested in the brain *qua* representer, in which case she might ask about the mapping that takes neural elements into a represented domain. Each project requires a mapping between different types of domain: in the former case from phenomenal to brain states, and in the latter from representations to represent*eds.* Köhler's principle interest was in the mapping from phenomenal to brain states. I'm not entirely sure which mapping SHEPARD is principally interested in, despite his usual talk about "representation."

In each case theorists have used the idea that the neural domain is isomorphic to what we might call the target domain (be that a set of representations or a set of phenomenal states). The idea behind such theories is that the domains *share* a structure. This identity of structure allows us to learn about the neural structure by studying the target structure. The most simplistic description of the Gestalt case has it that when we find a circular percept in our "behavioral environment" (Koffka 1935), we can infer that "concomitant physiological processes" (Köhler 1929, p. 61) share this circular structure. And while few theorists today hold that there is such a straightforward mapping, the very same principle lies behind the idea that the psychological color space with its three axes maps onto a neural coordinate space with sets of opponent process cells realizing each axis (e.g., Hurvich 1981).

The same problem haunts every version of such a theory, but it is most easily seen in the spatial case. We know that the sense in which the neural domain has a circular structure will differ from the sense in which the phenomenal domain has that structure – the neural version won't look circular to an observer gazing at the cortex. There *is* a trivial sense in which it has a circular structure just if it realizes a circular percept, but I take it a theorist interested in the nature of the brain would want a better description, like a map of the surface of the cortex. Yet, however we characterize the structure shared by the phenomenal and the neural domain, it must be abstract enough to describe both domains; so it cannot be written in neural terms. Certainly **SHEPARD**'s are not. The case is the same with structures characterized otherwise: if the terms in a functional characterization (a type of structure) do not refer "transparently" to entities and relations in the domain, then we need a translation into the particulars of a domain. Some (e.g., Fodor) maintain that psychological laws are irreducible, hence that no translation is required. But I doubt that SHEPARD is of this school.

**SHEPARD** has famously invoked isomorphism as well. He writes that "the default motions that are experienced in the absence of external support are just the ones that reveal, in their most pristine form, the internalized kinematics of the mind and, hence, provide for the possibility of an invariant psychological law" (target article, sect. 1.7). These laws do not directly govern brains, any more than the "next-to" relation in the visual field tells us about the next-to relation in cortical cells. Such judgments are reports about the structure of a state space of subjective states, a structure shared by the brain. Shepard has devised ingenious techniques whereby we can learn about the structures of psychological domains, and his present paper describes several. The large-scale project is the same as Köhler's: a set of laws that describe the behavior of phenomenal states (as those in accordance with Chasles' theorem) describes a *structure* of those states. In contrast to Köhler's phenomenal circle, this structure is distributed over time and counterfactual situations. But we must still say how this particular structure is realized in neural stuff. If the realization is abstract, how does this characterization constrain the possible configurations of the neural realizers? How are we to arrive at a translation from the characterization of the structure SHEPARD proposes to the description of a brain?

An important element in **SHEPARD**'s argument is that selective pressures favor the internalization of law-governed regularities. He writes of "the benefits of representing objects *as* enduring entities" (emphasis mine). The most straightforward way such internalization might proceed is to internalize a domain that directly realizes the laws that constitute the structure. But this is as implausible as the direct realization of phenomenal structures; it is more likely the case that we simulate the structure indirectly. But here the distinctions between various domains are important. We are out to find an informative description of the brain that tells us either how it represents or how it realizes phenomenal states. But the mapping from *what* is represented to *how* it is represented (i.e., to the nature of the vehicle) is no more straightforward than



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

the mapping from phenomenal states to brains. To represent movement is not to have a moving representation – presumably we represent without using the very properties that are represented. But then why should we assume in this case that representations of a certain *type* of movement move in the very *way* that is represented? We are left with a pair of troubling facts: knowing the structure of one's phenomenal states does not seem to tell us enough detail about the structure of one's brain, and knowing the structure of a represented domain does not seem to entail anything about the brain either. Suppose *both* that Chasles' theorem governs phenomenal states, *and* that we were selected to represent objects in the world as obeying Chasles' theorem. What follows about brains?

## Dynamics, not kinematics, is an adequate basis for perception

Andrew Wilson and Geoffrey P. Bingham

*Department of Psychology, Indiana University, Bloomington, IN 47401.*
**{anwilson; gbingham}@indiana.edu**

**Abstract:** Roger Shepard's description of an abstract representational space defined by landmark objects and kinematic transformations between them fails to successfully capture the essence of the perceptual tasks he expects of it, such as object recognition. Ultimately, objects are recognized in the context of events. The dynamic nature of events is what determines the perceived kinematic behavior, and it is at the level of dynamics that events can be classified as types.
**[SHEPARD]**

Roger **SHEPARD** has produced a fascinating account of how one might go about functionally representing the world. However, he has failed to successfully motivate his account. His evolutionary motivation is transformable into a story in which the advantage lies in the organism being able to flexibly perceive the world as presented to it (to support functionally effective action), rather than perceive the world represented by it. Also, it is not clear how conception should be uniquely related to perception and, therefore, how the study of conceptual problem solving can tell us anything about perception. SHEPARD uses his findings from studies on thinking to make claims about the nature of perception. We take issue with these claims.

**SHEPARD** notes, "My own searches for universal psychological principles for diverse perceptual-cognitive domains have been unified by the idea that invariance can be expected to emerge only when such principles are framed with respect to the appropriate *representational space* for each domain" (Introduction, pp. 581–82). His space is defined by canonical "landmarks" and the kinematic transformations that can transform them to match what is being perceived. We agree with his sentiment, but will argue that discrete kinematic transformations on time-independent landmarks do not make a suitable space. The space must contain continuous spatial-temporal forms because these (dynamically determined) "event forms" allow events to be identified as an example of a type, enabling recognition *of* the event and of objects *in* the event. A smooth gray surface could equally well be on a ball made of styrofoam or rubber. The continuous forms of motion exhibited when the objects are dropped on the floor allow them to be recognized for what they are.

Any motion, according to **SHEPARD**, is compatible with a dynamical account, given *arbitrary unseen* forces in nature. In his research where two discrete images are presented serially to subjects, they react in a manner consistent with a mental transformation of the initial object into the second via the kinematically optimal path. Not assuming the "arbitrary" forces makes kinematics the more empirically adequate account of these results. SHEPARD uses this as evidence that kinematics is the internalized transformation rule, and bolsters this claim with the additional claim that only kinematics, and not dynamics, is visually specified and, therefore, available for internalization. If dynamic properties, such as mass distribution and its consequent inertial properties, are not uniquely specified visually, they cannot serve as a reliable, internalizable set of transformation rules.

An error implicit in this argument is the suggestion that the forces in nature are arbitrary. They are not. Dynamicists study and describe reliable regularities in nature, configurations of force laws that produce the invariant forms of events that make the event recognizable. **SHEPARD**'s mistake is to exclude these consequent forms as "unseen." The effect of gravity on motion is perceptually salient, and observers are competent in using this information to identify events. The dynamics are immanent in the recognizable event whose spatial-temporal form is determined by those dynamics. That is not to say that gravity per se is necessarily recognized for its specific role in constraining the motion of a bouncing ball. Pre-Socratic philosophers need not have recognized gravity, but they certainly would have recognized bouncing balls, and observers fail to recognize bouncing balls as such when gravity has been altered (Twardy 1999; submitted). Events are recognizable according to the invariant kinematics, which, in turn, is determined by the underlying dynamics (Bingham 1995; 2000; Bingham et al. 1995; McConnell et al. 1998; Muchisky & Bingham, in press).

Two fundamental attributes of representational accounts are the ability of the representation to become causally de-coupled from that which it represents, and the low-dimensional nature of any computationally feasible representation. In **SHEPARD**'s account, these are achieved by reducing the visual space to a manifold of that space, defined by templates and geometric transformations for restoring something like the full dimensionality of our perceptual experience. In order to generate a veridical perceptual experience, the representation begins with its templates (and so must be in an appropriate initial position in the representational space), and transforms that template according to its geometrical rules to match the object being perceived. For this to work, both steps are highly constrained by the details of the event itself. If they are not, then the two attributes noted above mean that the veridicality of the perception is in doubt. In order for a representation to serve as the basis for tasks such as object identification it must be capable of generating constant matches and rematches between actual objects in motion and represented-object-transformed-by-geometrical-rules.

Imagine viewing a leaf blown about on a gusty day. By **SHEPARD**'s account, the perception of this leaf would require the computation and constant re-computation of the match of the leaf to a canonical "leaf." This computation would filter out the kinematic details specific to gravity and aerodynamics. Now imagine the perception of dozens of leaves on a gusty fall day. The representational structures would have to be generating constant, real-time updates for each of the large number of structurally similar objects all moving very differently in terms of the momentary directions and orientations (but moving in the same fashion in terms of the dynamically determined type of event). In Shepard's account, the computations would be defined kinematically, whereas what we actually see is best defined *dynamically.* This means that an unconstrained reconstruction could easily fail to use the appropriate kinematic mapping for the event's dynamic properties.

The reason for this can be thought of this way: kinematics is a description of a particular, local motion, one specific trajectory through a state space. Dynamics is the abstract description of the state space itself, describing the entire set of possible motions that correspond to a type of event. A pendulum can exhibit very different kinematic behavior. It can spin in circles, or swing in an arc. Both of these are captured in a single dynamical description, and it is only at the level of dynamics that these two motions can be classified as the same type of event, namely a pendulum event (Bingham 1995). Similarly, on a fall day, one does not merely see a collection of leaves; one sees a lot of falling wind-blown leaves (as compared to leaves that are merely falling on a calm day).



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

**SHEPARD**'s representational model is clear and well formalized, and has made conceptual contributions to recent work in computer vision (Edelman 1999). However, the model is demonstrably not about human perception. Edelman's implementation is interesting for object recognition in image based processing. Human vision is not based on static images, however, and experiments implying kinematic reconstructions of potential transformations between discrete images are not a fair test of the human visual system's event perception capabilities. Using such experiments to separate the role of the perceiver from that of the dynamic world removes nearly all of the relational information normally available to, and utilized by, the perceiver in event perception, and changes the nature of the task. The world is such that the nature of the perceived event is specified adequately while the organism is causally coupled to that event. When not coupled to the event, things such as imagination, dreams, or imagery are, as **SHEPARD** claims, likely to be derivations based on the way the event was originally perceived. If kinematics is insufficient for the original perception, it is therefore unlikely to be sufficient for any related task.

## Internalized constraints may function as an emulator

Margaret Wilson

*Department of Psychology, North Dakota State University, Fargo, ND 58105.*
**margaret_wilson@ndsu.nodak.edu**
**http://www.melange.psych.ndsu.nodak.edu**

**Abstract:** Kubovy and Epstein's main quarrel is with the concept of "internalization." I argue that they underestimate the aptness of this metaphor. In particular, an emulator which predicts unfolding events can be described as an internalization of external structures. Further, an emulator may use motoric as well as perceptual resources, which lends support to Hecht's proposal.
**[HECHT; KUBOVY & EPSTEIN]**

Has evolution caused our perceputal systems to "internalize" the constraints of the physical world? **KUBOVY & EPSTEIN (K&E)** answer this in the negative, but more on grounds of phrasing than on grounds of fact. To address their quarrel with this question, it may be useful to divide it into three questions:

(1) What abilities or tendencies have been built into our perceptual machinery? Do they mirror the constraints of the physical world?

(2) Are these tendencies represented explicitly in the form of rules, or are they present implicitly in the functioning of the perceptual machinery?

(3) Do these tendencies deserve to be called internalized constraints? Is any explanatory power added by conceptualizing things in this way?

**K&E** appear to agree with **SHEPARD** on the first question. They insist that constraints are something out in the world, not in the head; but they also acknowledge that to benefit from these constraints the perceptual system must have been shaped to produce output concordant with the constraints.

It is with the second question that the quarrel begins. **K&E** read **SHEPARD** as favoring the explicit view, with "mental contents actively engaged in the perceptual process." In contrast, they themselves prefer an account in which the system does not *follow* rules, but rather *instantiates* them.

How much, though, really hangs on this question? It is surely an interesting question in and of itself (though a notoriously difficult one to answer with any satisfactory clarity); but does the answer have substantial consequences for **SHEPARD**'s idea of internalization? **K&E** concede that "the visual system proceeds *as if* it possessed knowledge of kinematic geometry." But if there is fundamental agreement on the behavior of the visual system, and on how closely it tracks the properties of the physical world, it seems to me that Shepard's point is carried. The precise nature of the machinery which does this tracking can be left to future investigation. (The importance assigned to this question, though, may come down to personal preference. Some investigators, such as myself, have considerable tolerance for the "as if" style of explanation so pervasive in cognitive psychology, while others find it unacceptable.)

The quarrel continues with the third question. **K&E** object to the word "internalization," and suggest that it is an appealing metaphor but ultimately not an appropriate one. Yet, while they suggest a number of reasons why such theoretical terms *can* be problematic, they don't clearly state an objection to this particular term. I suspect that their objection in fact rests on their answer to the second question. The "internalization" phraseology suggests that the constraints of the external world are "things" that have been lodged in the mind in the form of explicitly stated rules.

But the internalization metaphor need not be read this way. External constraints – the laws of physics and so on – are not themselves explicitly represented rules. Instead, they are implicitly instantiated in how the system, the physical world, behaves. And in response to evolutionary pressure, based on the realities of living in such a world, a parallel set of "constraints" has sprung up within the cognitive system, again instantiated in how the system behaves. In such a situation, where external constraints cause the emergence of a mirroring set of internal "constraints," it seems to me that "internalization" is a particularly apt metaphor. (**K&E** make reference to the work of Lakoff & Johnson 1990 regarding the metaphorical basis of abstract concepts, but neglect the point that these metaphors are so pervasive, systematic, and enduring because they often track reality so well.)

Is any useful purposed served, though, by this "internalization" terminology? I suggest that there is, not least because it invites contact with a related set of ideas gaining currency in the field. A problem faced by any physical system that must interact with the world in real time, be it a remote-controlled factory robot or a human body, is that of making corrections and adjustments in response to feedback that is delayed by the time required for signal transmission. Even a slight delay in feedback can result in overcompensation for errors, which then require further compensation, and so on. One possible engineering solution is the use of an "emulator," a mechanism within the control system that mimics the behavior of the situation being acted upon, taking afferent copies of motor commands and producing predictions of what should happen (e.g., Clark & Grush 1999; Grush 1995).

To be useful, an emulator of course needs to be a successful mimic of the external situation. And to be a successful mimic, it almost certainly needs to be structurally isomorphic, for certain relevant properties, to the situation being emulated. Although much that happens in the world is unpredictable, there is also a good deal of regularity and redundancy that could be exploited by an emulator. It does not seem far fetched, in such circumstances, to say that the structure of the situation has been internalized. This may be just what has occurred in the evolution of the human visual system.

If the human visual system does indeed use this kind of an emulator, then the percepts we experience under conditions of ambiguous or "gappy" input would presumably reflect the predictions produced by the emulator. This may help to bridge the gulf between the minimalist stimuli of the laboratory and the more robust input usually provided by ordinary perception. The emulator functions in both situations, but is only allowed to truly determine the content of the percept when there is a temporary absence of reliable input.

Further, it is possible that this same emulator, run off-line, is responsible for the phenomenology of mental imagery (cf. Grush 1995). If this is true, and if there is indeed an isomorphism between emulator and world on some restricted set of relevant properties, then "mental rotation" may in fact be a mental process that is isomorphic to the actual rotation of a physical object.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

The concept of an emulator may also provide a boost to a suggestion offered, but only weakly supported, by **HECHT**. I have argued elsewhere (Wilson, in press) that there is a class of stimuli for which the outputs of the emulator will be particularly robust. These are stimuli that are "imitatable," in the sense that the posture or movement of one's own body can be mapped onto the posture or movement of the stimulus. The most common stimuli are of course the bodies of other people, but other stimuli, such as dogs and flying baseballs, can be imperfectly mapped onto all or part of one's body. In emulating an imitatable stimulus, not only perceptual constraints (I use "constraints" here to mean the internal principles that mirror the external ones) but also motoric constraints can be brought to bear.

**HECHT**'s proposal that, for certain perceptual judgments, we "project" our motoric knowledge onto inanimate objects – for example, judging the acceleration of a ball by unconsciously considering how a hand would accelerate to throw a ball – is clearly similar in spirit. HECHT uses an unfortunate choice of words in calling this "externalization," apparently in an attempt to emphasize its opposition to **SHEPARD**'s "internalization." But the two ideas are not mutually exclusive, let alone mutually exclusive *and* exhaustive as HECHT seems to suggest.

Indeed, **HECHT** weakens his own proposal by attempting to undermine **SHEPARD**'s. Fortunately, the attempt does not succeed. HECHT offers two counterexamples to SHEPARD's principle, both taken from a domain that Shepard has already identified as an unlikely candidate. At the same time, HECHT declines to discuss other more promising candidates, such as the Gestalt principle of common fate, on the rather odd grounds that the Gestaltists themselves were not interested in internalization. His argument that internalization must occur in all possible cases, or suffer severe diminution of importance, is not persuasive.

This is just as well, since **HECHT**'s proposal of motoric constraints is much stronger within a context where perceptual constraints are also operating. Motoric constraints alone would be of limited value and applicability. But within the context of an emulator, which might be expected to exploit regularities from any available source, motoric constraints might well be a valuable means of supplementing our ability to predict the unfolding behavior of the situation in a rapid, on-line fashion.

## Why perception is veridical

Alf C. Zimmer

*Department of Experimental Psychology , University of Regensburg, D93040 Regensburg, Germany.* **alf.zimmer@psychologie.uni-regensburg.de**
**http://www.zimmer.psychologie.uni-regensburg.de/**

**Abstract:** In pointing out that space and time are a priori constraints on perception and cognition, Immanuel Kant in his *Critique of Pure Reason* (1781) left open the question how perception is coordinated with the transphenomenal reality (the "Ding an sich"). I argue that the solution to this epistemological core problem of psychology is implicit in Roger Shepard's concept of psycho-physical complementarity.
**[SHEPARD]**

The real bone of contention Kant threw to psychology was not the disparaging remark that psychology was not a science (in his *Metaphysical Foundations of Natural Science,* 1786/1970) but the question (in his *Critique of Pure Reason,* 1781/1968): How can perception be veridical if the transphenomenal reality (the "Ding an sich") is categorically different from the phenomenal world we perceive? Since then, a strain of constructivist theories of perception has developed in psychology and biology, starting with Helmholtz (see especially his theory of "unconscious inferences"; Helmholtz 1856/1962) and leading to Maturana (see exposition of his position in Lettvin et al. 1959) and Rock (1983) – to name a few of the especially influential scientists of perception. Despite all the differences in detail, the common feature of constructivist theories of perception is the clear distinction between the transphenomenal world (*reality,* i.e., the world of things) and the phenomenal world (*actuality,* i.e., the world we interact with). Perception only relates to the latter and therefore mechanisms of perception (e.g., Kant's a priori of space and time) do not reflect upon the constraints of the physical world (reality). Gestalt psychologists, like Köhler (1958) or Koffka (1935), have tried to bridge the gap by taking the epistemological stance of "critical realism" according to which the phenomenal world is tuned to reality by experience – Köhler, by postulating psycho-physical parallelism and, ultimately, the isomorphism between percepts and brain states; Koffka, by stating that only those mechanisms of perception have evolved which resulted in a better inclusive fitness of the organism applying them. However, the question remains open, whether our perception shapes the world or the world shapes our perception.

In 1981, exactly 200 years after the first publication of *The Critique of Pure Reason,* Roger **SHEPARD** put forward the concept of psycho-physical complementarity in which he answered the question which had troubled Gestalt psychologists: "(1) The world appears the way it does because we are the way we are; and (2) we are the way we are because we have evolved in a world that is the way it is. In short, we 'project' our own inner structure back into the world, but because that structure has evolved in a complementary relation to the structure of the world, the projection mostly fits" (Shepard 1981b; see Kubovy & Pomerantz 1981, p. 332). His theoretical, as well as his experimental work since then has amassed arguments in favor of this position. Even if we are still far away from the final solution of this central epistemological problem, the system-theoretic solution implicit in Shepard's approach seems to provide the tools for attacking this problem successfully; it should be noted that W. Köhler in 1927 already toyed with the tools of systems theory but failed to apply them to perception.

The question remains, why has **SHEPARD**'s unique blend of evolutionary and systems theory not won the field in the battle over the veridicality of perception? My tentative answer is that the 2,500 years of philosophers repeating over and over again the metaphor of "nature as a book we read" has molded our thinking so strongly that in analyzing perception, theoretical concepts like structure, logic, or grammar simply "pop up" first and then govern the subsequent theorizing successfully. Consequently, in this line of thinking the veridicality of perception does not play a role.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

# Authors' Responses

## Shepard's Response

## On the possibility of universal mental laws: A reply to my critics

Roger N. Shepard

*Department of Psychology, Stanford University, Stanford, CA 94305-2130.*
**roger@psych.stanford.edu**

**Abstract:** In psychology, as in physics, principles approach universality only if formulated at a sufficiently abstract level. Among the most fundamental principles are those of generalization and inductive inference and those of perceptual and mental transformation. With respect to their appropriate abstract representational space, the former principles are well formulated as (Bayesian) integration over suitable (e.g., connected) subsets of points in the space, and the latter as geodesic (hence, least-time) paths between points in the space. Critics sometimes insufficiently appreciate the following: (a) Generality requires such abstraction. (b) Perceptual principles are not themselves given in sensory input. (c) Principles of learning are not themselves learned. (d) Though all such principles are somehow instantiated in the brain, their ultimate, nonarbitrary source must be sought in the regularities of the world – including those reflecting abstract mathematical principles (e.g., of group theory and symmetry). New light may thus be shed on the cognitive grounds of science and ethics.

### SR0. Introduction

I would be foolish indeed to take rejection of my ideas as confirmation of their validity. Yet, I find some comfort in the thought that immediate and universal acceptance of those ideas would have raised the question of whether they are sufficiently novel. I only hope that the majority of my critics might heed Mark Twain's counsel: "Whenever you find you are on the side of the majority, it is time to pause and reflect."

Hence, I begin this Response by elucidating, in the following overview section SR1, some of the general considerations that have led me to the evidently controversial approach to psychological science represented by my target article (reprinted from Shepard 1994). Then I respond, in each of the ensuing sections SR2, SR3, and SR4, to the articles and commentaries primarily concerned with the three topics of the representation of motion, color, and probability of consequence, respectively. In section SR5, I reply to the commentaries concerned with more general aspects of my whole approach. Within each of these sections SR2 to SR5, I take up the individual commentaries in an order that, I hope, serves to be coherent and to clarify my own views.

In the concluding section SR6, I offer a short summary of how, since the original 1994 publication of my target article, I have been seeking to extend the approach it represents toward two issues of broader significance. These concern the cognitive grounds, first, of science and, second, of ethics.

### SR1. General overview

#### *SR1.1. The pursuit of "Why" questions*

A case could be made that investigations at the more primitive, intermediate, and advanced stages of scientific development are primarily motivated, respectively, by "What," "How," and "Why" questions. The psychologist or ethologist might start by asking *what* patterns of human or other animal behavior occur in particular situations. The neuroscientist might then ask *how* these patterns of behavior are generated and controlled by mechanisms within the individual. And the evolutionarily oriented cognitive theorist might ask *why* these patterns have the particular form that they do. Notice, here, that the neuroscientist's most ready answer to *this* "Why" question – namely, the answer that the behavioral patterns are determined by the underlying neural mechanisms – only begs the further question of *why* those neural mechanisms are the way they are.

Ultimately, answers to such "Why" questions of psychological science can be found only through consideration of the world that has selected and then fine-tuned the individuals of interest. Thus, although some commentators attribute to me a more one-sided position, I hold that such shaping comes both through the eons of preceding natural selection of an individual's ancestral line and through that individual's own interaction with the world – that is, through both evolution *and* learning.

I do, however, regard the evolutionary shaping as more fundamental, for several reasons. It has operated for an incomparably greater period of time. The principles of learning – not having themselves been learned – can only have been shaped through natural selection. As analyses by machine learning theorists have rigorously confirmed, there can be no learning by a system lacking principles of learning sufficiently matched to the world in which that system is to learn (e.g., Hausler 1988; Mitchell 1980/1990; Watanabe 1985; Wolpert 1994; 1995). Equivalently, as I have long emphasized, because an individual rarely, if ever, reencounters the identical total situation, learning is of no use unless the individual is already endowed with a metric of similarity governing how (in the world!) that learning should generalize to newly encountered situations. And finally, while some organisms arise, survive, and propagate without benefit of any significant capacity for learning, in the absence of natural selection, no organisms arise, survive, propagate, *or* learn.

People tend to equate intelligence with the capacity for learning and, accordingly, denigrate those animals that act by "mere instinct." But it is precisely the most transcendent intelligence – an omniscient god or Laplacian demon – that already knowing everything about all things at all places and all times, would derive no benefit from a capability for learning. In our own present stage of internalization of knowledge about the world, we fall far short of such transcendent intelligence and, hence, a capacity for learning is of enormous value.

Nevertheless, each of us comes into the world with deeply embodied innate wisdom about some pervasive, enduring, and biologically relevant features of our world. Though generally inaccessible to conscious introspection, this implicit wisdom eliminates the risk of our having to learn some biologically significant facts about the world, as I have put it, "by trial and possibly fatal error." It also fur-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

nishes us with principles – attuned to this world – of individual learning about other circumstances of comparable significance that are more local, transitory, and contingent. In addition (as I will discuss in the concluding section SR6 of this Response), there is much knowledge that we humans *can* acquire, through science, about other features of the world that are universal, but on scales of space, time, or energy that have been of insufficiently direct relevance to our ancestors to have been in any way internalized.

Several commentators complain that my ("Why"-motivated) quest for universal psychological laws neglects their own ("How"-motivated) desire for explication of the brain mechanisms which they regard as the source of any such laws. I offer two reasons for my strategy. First, given the unprecedented complexity of the brain, there is little hope of comprehending its modes of operation without first understanding the problems it has evolved to solve. The advances in our understanding of neuronal mechanisms (as well as in devising robotic control mechanisms) indicates that those advances were almost always preceded and guided by knowledge of the problems that such systems confront in the world. (I found telling, in this connection, this confession in the article by the neurophysiologically-oriented vision researcher Horace **BARLOW**: "If I had been smart enough I would have predicted the orientational selectivity of cortical neurons that Hubel and Wiesel . . . discovered . . . ")

Second (and more fundamentally, from my standpoint), selection by the world – whether through evolution or through learning – is directly contingent on the overt behavior of individuals in the world. Such selection does not discriminate among alternative mechanisms capable of generating that same behavior. (This is analogous to the case of vision, where selection favors the emergence of an image-forming organ. But such selection does not dictate whether that organ consists of a single lens and an array of photo-sensitive cells oriented away from the lens, as in the vertebrate's eye, or toward the lens, as in the cephalopod's eye; or of many hexagonal lenses, as in the insect's eye, or of square-cross-section wave guides, as in the eye of some crustaceans. **Vallortigara & Tommasi**, in their commentary, provide another possible example of such convergent evolution of function in which humans and birds, despite gross anatomical differences in brain structure, evidently resolve a type of visual ambiguity in the same way.) As Cosmides and Tooby have observed, "Absent a theory of function, there is no basis for deciding which machines are functionally analogous" (Cosmides & Tooby 1997, p. 154). In short, I suggest that universality is more likely for functions that arise as direct accommodations to continuing universal demands of the world, than for the various anatomical and physiological structures that have come to implement those functions through an evolutionary history of "frozen accidents."

According to the traditional hierarchy of the sciences, psychology is far separated from physics and mathematics by the intervening sciences of chemistry, biology, and neurophysiology. Yet, universal laws resembling those of physics and mathematics may be more widely achievable in psychology than in biology. The marked structural differences between land-dwelling mammals, cetaceans, bats, birds, snakes, fish, molusks, arthropods, and so on, attest to the diverse ecological niches through which their various evolutionary histories have branched. But all of these species, regardless of the peculiarities of their particular niches, must represent, alike, a world of semi-permanent objects belonging to distinct kinds and susceptible to motion in three-dimensional space.

No doubt there are some biological universals. The genetic code and the particular chirality of the terrestrial DNA helix are possible candidates, though whether these are true universals or consequences of some early frozen symmetry-breaking accidents in the molecular origin of life on Earth, may be uncertain. My one best candidate for a life-science universal is a more abstract evolutionary mechanics such as may be common to all systems that adapt through natural selection – regardless of the specific details of the molecular implementation of that evolutionary mechanics.

In any case, I suggest that despite the enormous diversity of biological forms, natural selection for effective function in the world favors the emergence of cognitive representations that reflect an ever widening class of features of the world. For an evolutionary line, such as that of Homo sapiens, which has preeminently branched into what has been termed the general "cognitive niche," mental laws may be selected to achieve an ever closer mesh with relevant physical and mathematical demands of the world. For such a line, I have suggested that the traditional linear hierarchy of the sciences might be said to be "curled" toward the closing of a "psychophysical circle."

### *SR1.2. Toward universal, non-arbitrary psychological laws*

To the extent that laws proposed within any branch of science – whether physical, biological, or psychological – are but descriptive summaries of data obtained for some narrow and possibly accidental set of circumstances, those "laws" will almost certainly be found deficient when tested under a wider range of circumstances. This was the fate of Aristotle's principle that an "impetus" imparted to a body (in the "sub-lunar" realm) decays until the body attains its natural state of rest on the Earth – a principle undoubtedly based on the familiar behavior of objects that have been given a push under the locally prevailing terrestrial conditions of surface friction and air resistance.

Likewise in biology or psychology, there would be little justification for putting forward, as a universal principle, a regularity found in the form or behavior of a species that happened to have branched into a restricted terrestrial niche. Principles with such a limited empirical basis may have practical utility for dealings with particular terrestrial animals, but they hardly qualify as universal laws of life or of mind.

Toward what kind of psychological science, then, shall we aspire? Physics and, certainly, mathematics can boast elegant laws that hold universally. In contrast, psychology (like biology) is widely regarded as limited to the discovery of principles pertaining to the particular organisms that have evolved by adapting to particular series of accidental circumstances on this particular planet. After all, those are the only organisms available to our empirical study.

Hoping for a grander science of mind, I have been exploring the implications of the fact that some of the circumstances to which we may have adapted are not accidental or peculiar to planet Earth. Some quite abstract yet biologically relevant circumstances are presumably characteristic of *any* environment capable of supporting the evolution of intelligent life. The universal circumstances on which I have focused are quite abstract: (a) Space, on a



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

biologically relevant scale, is three-dimensional and Euclidean. (b) Biologically significant (macroscopic) objects, while having six degrees of freedom of global position in this three-dimensional space, tend to maintain their integrity over extended periods of time – typically preserving mass, significant aspects of shape and of surface spectral reflectance characteristics. (c) Such biologically significant objects, if possessing sufficient mutual similarity, also tend to belong to some distinct basic *kind*, with its own inherent degree of potential either for furthering or for threatening an individual's wellbeing or reproduction. In my target article I tried to indicate how from such abstract features, natural selection would in the long run tend to favor individuals who have "internalized" principles of representation and behavior that approximate what would be optimum in a world with such features.

### SR1.3. General comments on the preceding articles and commentaries

Some authors reject my proposed universal principles as unfalsifiable. Others reject them as already falsified. People of these two persuasions should talk to each other! Yet it falls to me to have the last word. I can only suggest that both of these two seemingly incompatible types of criticism stem from misunderstandings of what I have proposed.

Most of the commentators who claim that the principles I have proposed have already been disconfirmed, argue that this or that concrete and specific aspect of the world is not internally represented. In doing so, they ignore my repeated declarations that it is the most pervasive and enduring features of the world that are likely to become most deeply internalized.

Many persist in assuming, for example, that what I am (or should be) proposing about the principles underlying apparent motion or mental rotation is that they must be based on the ways in which actual physical objects most typically move in the world. But the fact about the world from which I derived my proposed kinematic principles of spatial transformation is a much more abstract and universal one. Specifically, it is the fact that the possible positions of an object correspond to points in an abstract six-dimensional space in which the simplest (and, I conjecture, the quickest) confirmation of the shape-identity of two objects is achieved by the mental transformation that corresponds to a traversal of the shortest path in that space along a geodesic between the points corresponding to the two objects. (Such a traversal is "simplest" in the sense that it is generated by iteratively applying the identical very small spatial operation – Carlton & Shepard 1990; Goebel 1990.)

Similarly, concerning my theory of generalization, some commentators mistakenly take my derivation of an exponential law to be based on an assumption that the "basic kinds" of objects, events, or situations that are biologically relevant for a particular species at the present time are universal. But I derived the proposed universal law from the much more abstract and universal fact about the world that biologically significant objects belong to basic kinds – without assuming anything whatsoever about what specific kinds are prevalent or relevant for any particular species at any particular time.

Failing to grasp the fundamental tenets of my general approach, several commentators question the strict validity or universality of the subsidiary or provisional specifications I sometimes make to facilitate mathematical analysis or empirical test. In doing so, they suppose they have established the untenability of my whole approach. But the general principles that I derive can be shown to be quite insensitive to the specifics of these subsidiary or provisional specifications.

But what about the different worry that the principles I have proposed, far from having been falsified, are in fact unfalisfiable? I admit that some of what I have set forth might better be described either as a general framework for a whole class of formally related theories, or as one very general theory with some parameters left open for future specification – to accommodate, for example, alternative perceptual-cognitive mechanisms, representational spaces, or strategies that may have so far been internalized as adaptations to different, partial aspects of the world.

Two such partial aspects are the geometrical principles of kinematics, which specify which paths of rigid transformation are geometrically simplest, and the principles of physical dynamics, which specify how physical bodies will actually move under particular conditions. I have argued that the former, geometrical principles are more basic and should therefore be more "deeply" internalized. Nevertheless, apparent motion or imagined spatial transformations might be guided by internal representations of either the geometrical or physical principles, or by some compromise between them, depending on the circumstances of test and on the history of the species or individual tested. For such reasons, provision has been made for parametric interpolation between the geodesic or "least-energy" paths appropriate to these cases (Carlton & Shepard 1990a; 1990b).

Does the provision of such parametric interpolation render the theory unfalsifiable? Not really. Even with free variation of the parameter, the variety of geodesic paths that can be generated is extremely restricted (in fact, has measure zero) relative to the set of all possible paths. Moreover, the theory itself provides some indication of the kinds of circumstances that should bias the system toward the representation of the geometrically simplest, or physically most natural, motion.

A related worry, for those who follow Gibson's approach, concerns the issue of ecological validity. What relevance does an account of apparent motion in terms of kinematic geometry have for the perception of actual motion when, as I myself acknowledge, one seldom (if ever) encounters instances in the natural world in which one static object disappears and another similar static object suddenly appears in a different location or orientation? My answer is that the purpose of the studies of apparent motion is not primarily to discover principles of motion perception, as such, but principles with more far-reaching implications. One implication is the existence of a spatial-transformational mechanism capable of quickly establishing the identity of objects despite differences in their orientations. Another is that this same spatial-transformational mechanism underlies our capabilities for problem solving, planning, invention, and perhaps (as I shall note in my concluding sect. SR6) the discovery of fundamental laws of physics through thought experiments.

The single objection that is perhaps most frequently raised by my critics is against my claim that knowledge about the world is internally represented or that enduring features of the world have already been "internalized." **KUBOVY & EPSTEIN** regard the notion of internalization as a useless or even misleading metaphor and many other commentators, following Gibson, hold that if the relevant information is


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

universally available in the environment, there is no need for it to be internally represented. In the words of **Jacobs et al.**, environmental "constraints are taken advantage of by detection of information granted by the constraints rather than by internalizing them." **BARLOW**, Robert **SCHWARTZ**, **TODOROVIČ**, **KUBOVY & EPSTEIN**, as well as **Jacobs et al.** suggest that internalization, thus failing to confer any advantage, would not be favored by natural selection.

In response to this objection, I note, first, that the relevant environmental information is not always immediately available to the perceiver and, second, that even when it is, to say that that information is simply "picked up," "resonated to," or "taken advantage of" is to ignore the fact that there must be *some* structure in the perceiver that enables it to pick up, resonate to, or take advantage of that information. To that extent, is it not appropriate to say that that internal structure "represents" something about the external situation that is "represented" by that information? A simple hypothetical example may help to clarify what I mean by "internal representation." It may also help to indicate how internalized principles can indeed confer a benefit by minimizing time, thereby optimizing the probability of a favorable outcome (much as I have argued for the principles underlying apparent motion and mental rotation).

A lifeguard, in attempting to rescue a drowning swimmer, must traverse a connecting path, from her lookout station set back on the wide beach, to the struggling swimmer way out in the water. Although her starting point and goal are given, her choice among the infinitely many possible paths from the one location to the other, though confined to the two-dimensional, approximately horizontal surface formed by the beach and the water, is in no other way externally constrained. But, of all these paths, the lifeguard wants to take the one that maximizes the probability of a favorable outcome, which (in this case) is the one that minimizes time.

If the least-time path were along the straight line between the lifeguard's station and the struggling swimmer, one might be tempted to say that the lifeguard could use the very simple strategy of both running and then swimming such that the drowning swimmer is always directly ahead of her. For this, it might be suggested, she needs no internal representation. (But even in this counterfactual case in which the least-time path is rectilinear, what is "directly ahead of her" is not represented in the same way in her "ambient optic array" when she is in the prone, swimming position, and periodically rotating her head to the side to breathe, as when she was running across the sand. Moreover, I would suggest, she must still in some way "represent" both her goal of minimizing time and the abstract principle that a straight line is the shortest and, hence, the quickest route to the drowning swimmer.)

The actual least-time path in the case considered is, however, not along the straight connecting line – except in the zero-probability circumstance in which that line happens to be exactly orthogonal to the water's edge. Even if the lifeguard is an excellent swimmer, she can run across the sand more swiftly than she can swim through the water. Consequently, to shorten the total time, she must run straight to a point along the water's edge that is farther from her than the point on the straight line between her and the swimmer. On reaching that point at the water's edge, she must change direction and swim along the thus shortened remaining straight line from this point to the swimmer. As noted by Feynman (1985, pp. 51–52), the resulting bend in the path is exactly analogous to the refraction of light as it crosses the boundary between a less dense medium (such as air) into a more dense medium (such as water) – a path deduced by Fermat from his teleological least-time principle for the propagation of light.

In the absence of external physical constraints, the path taken by the lifeguard is necessarily determined by her own internal processes and representations. Granted, achievement of the least-time path requires that these internal processes have access to, and properly make use of, the relevant facts about the physical situation. Two of these facts – the distance of her station back from the water, and the ratio between her top speeds of running and of swimming – may already be accurately internalized from her own prior experience. The other two critical facts – the direction and the distance of the struggling swimmer – however, must be perceptually gained at the time that she sights that swimmer. Following Gibson, the theorist seeking to describe the problem to be solved by the lifeguard would seek to identify the "higher-order" relation in the lifeguard's "ambient optic array" that (based on these relevant facts) "specifies" the direction in which the lifeguard should proceed to run across the sand. This higher-order geometrical relation exists but it is not simple. (My associates and I have yet to find a closed-form expression – though Thomas Griffiths came up with an efficient numerical approximation.) Merely to say that the lifeguard "takes advantage of" this complex higher-order relation ignores the internal structure that must be approximately tuned to this relation within the lifeguard.

Of course, the instantiated inner processes and representations that guide the lifeguard's choice of path need not be (and generally are not) anything like conscious symbol manipulations (such as a trigonometric calculation based on something corresponding to Snell's law governing the angles of incidence and refraction of light). They are generally of a more analogical character, largely inaccessible to conscious introspection or verbal report, and may only approximate the optimum solution (see my response to Peter **Todd** & Gerd **Gigerenzer** at the end of sect. SR5, and my closing sect. SR6). The extent to which these processes enable the lifeguard to approximate what is for her the unique least-time path is of course an empirical question. Yet, to the extent that they do, they might justifiably be said to have internalized an analog of Fermat's principle.

How, then, did Fermat himself arrive at the explicit formulation of his principle? We do not know; but his thinking may well have been guided by thought experiments (a topic I take up in my final sect. SR6). If so, the representations and principles that guided those thought experiments are not, surely, to be dismissed as the "detection of information granted by [co-present environmental] constraints" (of **Jacobs et al.**).

Finally, in proposing that any psychological principle is universal, I am not claiming, as some commentators seem to suppose, that the law in question explains all cognitive or behavioral phenomena, let alone such phenomena for all living species. Rather, I am suggesting that as life evolves toward greater perceptual and cognitive capacities, it should more and more closely conform to that law, under the simplified or purified test situations specified. This is analogous to the case in physics. Galileo's law of falling bodies is universal – not in that it, by itself, accurately describes the fall of every rock, leaf, and feather, but in that it is ap-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

proximated (anywhere in the universe) to the extent that friction, air resistance, and other extraneous factors are reduced to negligible levels.

I now respond more specifically (in sects. SR2, SR3, and SR4) to the individual authors who primarily focus on each of the three principal topics of my target article and, then, (in sect. SR5) to those who primarily critique my approach in general.

## SR2. On the representation of objects and their spatial transformations

### *SR2.1. Barlow*

Horace **BARLOW** shares with me the conviction that perceptual systems exploit regularities in nature or (as he puts it, for the case of vision) in "properties of natural images." Indeed, as he points out, we both share this general conviction with many of our predecessors, including Helmholtz and such subsequent theorists as Craik, Tolman, Attneave, and Brunswik. Although I am, accordingly, sympathetic with much of what Barlow says, there are some noteworthy differences in our viewpoints.

I myself credit Helmholtz with achieving perhaps the single most significant advance in our understanding of perception, with the formulation of his principle of *unconscious inference,* which I would express as follows:

The perceptual system generates an internal representation of the external object, event, or situation in a way that satisfies the following two conditions:

1. It is *consistent* with the information incident on the sensory surface.

2. It is, among all such consistent representations, the one corresponding to the external object, event, or situation *that is most probable in the world.*

The first condition is obviously necessary to avoid nonveridical perception and, hence, non-adaptive behavior. The second condition must be added because, for any given pattern on the sensory surface, there generally are many possible external situations that would yield that same pattern, and selection of the wrong representation would lead, again, to non-adaptive behavior.

I may differ from Helmholtz, in his most thoroughgoing empiricist stance, and from **BARLOW**, in his emphasis on learning, in the following respect. I claim that both conditions (1 and 2) of the above-stated principle depend on implicit knowledge that has, at least in part, been shaped by natural selection. Yet, if pressed, perhaps even Helmholtz and BARLOW would acknowledge that the structure of the perceptual system is to some extent already laid out at birth. Regardless of how much further shaping of the system is left to individual learning, an innate head start must confer some benefits on the individual.

Another respect in which I seem to differ, at least in emphasis, from **BARLOW** is (as he himself notes) in my belief that much of our knowledge about the world is more than merely "statistical." Although I speak (in Part 2 of my formulation of Helmholtz's principle) of what is most "probable" in the world, this is not really a statistical statement. The probabilities to be distinguished are often effectively either zero or one. The case of the retinal image produced by a distant object (such as a building) of cubical shape is illustrative. Such an image could have been produced by any of an infinite number of other shapes having wildly curved and non-parallel edges, as well as markedly non-orthogonal corners, provided only that each point of the shape falls on a straight line passing through corresponding points on a standard cube and on the retina.

One might be tempted to say that we experience a cube rather than any of these other, bizarre, non-cubical shapes (even though they all project identically on the two-dimensional retina) because cubes, or at least straight lines and right-angled corners, are statistically more prevalent in the world. But the fundamental reason for experiencing only the cube is less statistical: If the object had, in fact, been one of these other shapes (with their curved and non-parallel edges and non-orthogonal corners), the probability that a freely mobile observer would be viewing it from just that unique angle where the shape projects exactly the same retinal image as that of the highly symmetrical and regular, straight-edged and right-angled, standard cube is simply zero. (For further illustrations of this principle, see Shepard 1990b – especially pp. 48, 56, 133–136, 168–186; and Shepard 1981b, which cites earlier formulators of the principle.)

Similar considerations apply to the Gestalt phenomena of perceptual grouping. True, visual stimuli that are close together may tend to be perceptually grouped because things that are closer together in the world are somewhat more likely to be connected or be parts of the same object. But the Gestalt principle of grouping by common fate completely dominates the principle of grouping by proximity, when the two principles are pitted against each other. Why? Because the probability that objects that do not connect or communicate with each other in any way will just happen to simultaneously and in exactly the same way move (relative to their common background) is not just low – it is zero. Such considerations are, I think, less Brunswikian than Gibsonian.

The phenomena of apparent motion, first systematically studied by the Gestalt psychologists, are also well explained by Helmholtz's principle (as I have stated it). A circular dot presented to the left and then to the right, with no intervening movement or blank interval, would be equally consistent with either of the two following possible events in the external world. (1) A single, enduring circular object suddenly moved from the left to the right position (at too great a speed to have excited any retinal receptors along the path of motion). (2) A circular object suddenly went out of existence and, at the very same instant, a similar circular object materialized in a location to the right. In the actual world, however, material objects are generally conserved. Indeed, the probability is virtually zero for each of the following three components postulated for this second possibility: (a) a material object suddenly goes wholly out of existence, (b) a material object suddenly comes wholly into existence, (c) two spatially separated and unconnected material events occur simultaneously.

Other phenomena of apparent motion are also explained. If the inter-stimulus interval is lengthened, the experience of a single object moving back and forth becomes correspondingly weaker. Why? Because, if a real object took that long to move along the (straight) path between two positions, it would have moved slowly enough to have stimulated receptors along the retinal locus of the path. Since no such receptors were activated, such a motion is inconsistent with the available sensory information. Alternatively, the object could have moved over some much longer path and


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

at so great a speed that no receptors along that longer path were activated. But in this case, the perceptual system, having no basis for representing motion over any one such long (and possibly convoluted) path rather than another, would derive no benefit from hallucinating an arbitrary one of these possible paths.

(Apparent motion over a curved, longer path can however be experienced if a static, very faint simulated path of motion is presented during a brief inter-stimulus interval – as demonstrated by Shepard and Zare 1983. The stimulation of receptors along such a path *is* consistent with an actual motion of this kind.)

But, in the standard condition that yields compelling apparent motion – namely, one with very brief inter-stimulus intervals, and small spatial separations and low rates of alternation, – why is motion experienced over the straight connecting path? The primary answer, surely, is that rectilinear motion is, under these conditions, the simplest, least arbitrary, and most quickly generated. The capability of representing rectilinear motion may also be generally useful in that such motion approximates any smooth motion (especially of a circularly symmetric object) over sufficiently short intervals of time and distance.

Then why, in the modified condition reported by McBeath and Shepard (1989), in which symmetry was broken by removing equal but differently oriented pie-shaped sectors from the two alternately presented circular dots, is the motion experienced over a markedly curved path? And why, when two (asymmetric) three-dimensional objects of the same shape are alternately presented in two appropriately different positions, is a helical motion often experienced? Here again, I claim that kinematic geometry prescribes such motions as the simplest, least arbitrary, and most quickly generated. And here, over short intervals of time and distance, an arbitrary smooth motion is approximated (even more closely than by rectilinear translation) by a small screw displacement.

If the rate of alternation of the two stimuli is progressively increased, however, the experienced apparent motion abruptly changes – to a discontinuous, alternating appearance and disappearance in each of the two locations (without any connecting motion), in the case of the two circular dots. Why does this happen? And why does it change to a less curved path of motion or to a smaller experienced angle of rotation accompanied by a loss of experienced object rigidity under the conditions that otherwise induce rigid rotational motions?

Here, the distinction between competence and performance, drawn by Chomsky (1965) for linguistics is, I suggest, relevant for perception as well (see Shepard 1984, p. 429). That an individual's behavior falls short of what is prescribed by some theoretically optimum principle does not entail that the principle is not internally represented. The form in which the relevant external information is available may simply be too brief or sparse for perceptual assimilation, too extensive or fleeting for memory fixation, or too different from the ancestrally prevalent form to engage the evolved representational machinery.

Many of the phenomena of apparent motion would present problems for a theory that holds that such motion arose to subserve prediction by simulating the statistically most common motions of objects in the world. Several commentators seem to suppose that this is – or should be – my theory. But, as I have often noted, neither the paths of apparent motion nor the speeds of (mental) traversal over such paths simulate the paths or speeds of motion of physical objects in the world. Indeed, what would such paths or speeds be? (A bird may swoop or remain perched; a stone may lay still or be hurled.)

**BARLOW** (like several other commentators) suggests that my invoking of internalization "is insufficient because [for such a process to be favored by natural selection] it is also necessary to derive an advantage from the process." In reply, I propose that significant advantages are in fact conferred by the mechanism underlying apparent motion, and by the "internalization" of kinematic geometry. One advantage is that they permit the swift determination that successive sensory patterns correspond to the same, enduring external object. Apparent motion is the experiential accompaniment of the representational systems' establishing this correspondence. The motion may be experienced over a straight line, a circular arc, or a helical curve (as in the three cases referred to above). But this is not because they are the statistically most common motions of biologically significant physical objects. Rather, the inner enactment of these kinematically simplest connecting motions has been favored because there is a benefit in establishing the identification of the enduring object (and its change in position) in the shortest possible time.

The emergence of internal machinery capable of representing the simplest transformations between different positions of an object in space provides another significant advantage. This is the capacity for imagining the consequences of alternative possible actions in the world in advance of expending the effort, taking the time, or incurring the risk of performing those actions physically. My students and I have demonstrated this capacity through our experiments on "mental rotation" (Shepard & Cooper 1982; Shepard & Metzler 1971). These mental transformations, too, are characteristically performed not as simulations of the paths, speeds, physical likelihoods, or energetic feasibilities of any physical operations in the world. Rather, they are performed to establish a geometrical correspondence, and to do so as quickly as possible. (Such mental transformations may, however, also be controllable in a way that aids extrapolative prediction of actual physical motions or estimation of the efforts required to carry out alternative operations physically – as contemplated by Craik 1943).

Such a transformational capability is surely of practical utility – as in deciding whether a certain object will fit in a certain niche, or through a certain opening; or (as tested in the mental rotation experiments) deciding whether what has appeared in two different glimpses is, despite differences in orientation, of the same shape and, hence, may be the same external object. Moreover, (as I discuss in my concluding sect. SR6), such a capability may have made possible the "thought experiments" through which Archimedes, Galileo, Newton, Einstein, and others arrived at their fundamental physical laws. If so, the capability for such thought experiments would be a development of major significance for science and for the future of human kind that was not directly selected for in our ancestral line. Instead, it arose as a by-product of selection for the more immediate, practical benefits of the capability of representing concrete spatial transformations.

Despite the evident differences, **BARLOW**'s view and mine appear to have significant overlaps. For example, the importance that Barlow assigns to a rigidity principle and to



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

the hierarchical structure of perceptual processing accords with my apparent-motion demonstrations of a "hierarchy of criteria of object identity." As the rate of alternation is increased between two widely different orientations of the same shape, at some point an abrupt shift of the character of the apparent motion occurs. The previously experienced rigid rotation through the large angle is replaced by the experience of a non-rigid rotation through a smaller angle. Evidently, when there is insufficient time to represent the successive orientations of the complex object throughout the whole rotation, the system resorts to a weaker criterion of object identity. (An analogous phenomenon occurs in apparent motion of the human body – see Shiffrar & Freyd 1990.) It could be said (borrowing Herbert Simon's term) to "satisfice," by representing rotation through a smaller angle that achieves only approximate similarity of shape and, hence, yields an experience of non-rigidity (Farrell & Shepard 1981; Shepard 1981b). Depending on the shape of the object (specifically, its approximations to various symmetries), there can be several such abrupt shifts, each yielding the experience of a smaller rotation together with a larger non-rigid deformation.

Also, my proposal that kinematic geometry is to some extent internally represented, is in no way incompatible with **BARLOW**'s suggestion that "cortical neurons might usefully be selective for twisting motions." (There may even be some neurophysiological evidence for such a proposal – see, e.g., Gallant et al. 1993.) For, while the ongoing motions of biologically significant objects are not characteristically simple screw motions, any smooth rigid motion is, during sufficiently brief intervals of time, approximated by a small screw displacement (including the limiting cases of a purely translational or a purely rotational displacement). Neurons tuned to particular twisting motions might even help to answer BARLOW's question about how the system is so quickly able to select a candidate direction of motion in attempting to attain full congruence.

Less clear is how the existence of neurons that passively respond to particular twists would explain the active generation of successively more rotated representations of the structure of the whole object at successive orientations throughout a large angle of rotation. Yet, compelling evidence for such representations has been obtained in the cases both of mental rotation (e.g., Cooper 1975; 1976; Cooper & Shepard 1973) and of apparent motion (e.g., Farrell & Shepard 1981; Robins & Shepard 1977). It is nevertheless possible for large rotational transformations to be achieved by *actively* iterating very small ones within a parallel distributed processing architecture, as has been elegantly shown by Goebel (1990). Still less clear, is how passive twist-detecting neurons would contribute to ordinary mental planning of spatial operations, or to the physicist's or inventor's effective carrying out of thought experiments.

Finally, while I presume we all agree that some aspects of behavioral and mental capabilities are individually acquired and some are evolutionarily pre-wired, I readily admit that the empirical determination of the exact location of the dividing line is extremely difficult. Under these circumstances, I think it is healthy for psychological science that some theorists try to see how much can be accounted for by individual learning while others explore what might be accounted for by natural selection.

Justifications for pursuing the latter strategy include (once again) the following: (a) Much neuroanatomical structure is evident in the brain well before birth. (b) Core knowledge about some of the most basic physical, geometrical, and numerical facts is already present in early infancy (as demonstrated in the cognitive-developmental experiments of Elizabeth Spelke, René Baillargeon, Karen Wynn, and others – e.g., see Spelke 1991). (c) Studies in clinical neurology and behavioral genetics indicate that some individuals (with certain, apparently genetically determined conditions), despite exposure to the same opportunities for learning, suffer a specific cognitive deficit (such as the deficit in spatial problem-solving characteristic of Turner's syndrome). (d) The period of time that has been available for evolutionary shaping has vastly exceeded the time available for learning in each individual. (e) Individuals that come into the world already adapted to things that are generally true in the world must surely have some advantage over those that have to adapt to those things though trial and error. (f) Learning is governed by principles of learning and generalization that are not themselves learned but must be adapted to the world through natural selection.

Nevertheless, each of us must confront numerous specific biologically significant circumstances for which our evolutionary heritage cannot have prepared us. I defer fuller discussion of this problem to the later section SR4, on generalization.

### *SR2.2. R. Schwartz*

I appreciate the modest and inquiring spirit with which philosopher Robert **SCHWARTZ** seeks clarification of my proposal concerning "evolutionary internalized regularities." Because the questions that SCHWARTZ raises are very similar to several of those raised, at greater length, by **BARLOW**, I hope that my response to Barlow (together with my preceding introductory remarks), may provide at least some of the clarification that Schwartz seeks. I will, however, briefly comment on the following few specific concerns expressed by Schwartz.

"In Shepard's sense . . . the [internalized] constraint must be inherited or 'innate,' and not the result of learning" (**SCHWARTZ** target article, sect. 1, "Constraints and internalization"). And (in his next paragraph), "I am not sure what advantage Shepard's kinematic principle is supposed to confer." And yet again (in his concluding section), "If the actual movements our ancestors experienced were not by and large instances of the unique path specified by the constraint, what would drive or account for the evolutionary incorporation of the principle?" (sect. 6, "Evolution"). As I indicated in responding to **BARLOW**, I do not rule out internalization through learning. Yet, as I also indicated, a number of advantages are in fact conferred even by the evolutionary internalization of kinematic geometry.

**SCHWARTZ** objects that, "the conditions and stimuli used in the apparent motion experiments are not especially typical of normal movement perception. Hence, there is the worry that results found under these limited circumstances are not ecologically valid. They may not transfer or apply to cases of real motion . . . " (sect. 3, "The kinematic constraint and ecological validity"). Instead (he later concludes), "the constraint may only hinder perception of actual motions that do not fit its specifications. This . . . is especially troublesome, since much of the real motion we do encounter does not traverse a path that is the unique, twisting route prescribed by kinetic [sic] geometry" (sect. 6, "Evolution").


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Schwartz's "worry" is unfounded; natural selection has ensured (via the "consistency" part of Helmholtz's principle, as I stated it in responding to **BARLOW**) that sensory information about what is actually happening in the external world takes precedence over any default principles that come into play in the absence of such external information.

Even so, such default principles may actually assist rather than interfere with the perception of real motion. As I also noted in my response to **BARLOW**, the default principles of kinematic geometry do approximate what is going on, locally, in any smooth motion; and the kinds of twist-specific neurons that Barlow mentions may play some role both in the perception of real motion and, hence, may have contributed to the first steps of the original evolutionary implementation of Chasles's principle of kinematic geometry. My central point, however, is that the default principles have come to serve purposes quite different from that of inferring what physical motion is currently taking place in the external world. They serve, for example, the purpose of establishing that two sensory images are of the same external object.

"Unfortunately," **SCHWARTZ** adds, "the need to appeal to relatively non-normal conditions is in tension with a commitment to ecological validity" (sect. 3). My own position concerning ecological validity is a bit different. On one hand, in psychology (just as in physics), we often must set up experimental situations that significantly depart from those that are typical in nature, in order to eliminate complicating influences (analogous to friction or air resistance, in physics). Only thus can we, like Galileo, probe the pure principles operating behind the complexities of everyday events. In particular, if we want to discover any purely internal principles for the representation of spatial transformations, we have no choice but to devise a test in which no external motion is presented. On the other hand (as I later explain, in my response to **HECHT**), we must take care that our experimental stimuli preserve enough of the features of the situations that challenged our ancestors to enable these stimuli to engage the mechanisms that evolved to deal with those situations. There is indeed a tension here. But if scientific research came naturally and free of all tension, it would not have taken humankind so long to develop effective scientific methods.

### *SR2.3. Todorovič*

As I might appropriately do for many of the commentators, I thank Dejan **TODOROVIČ** for his initially positive characterization of my theorizing – as a "rare and welcome exception" to the usual neglect of evolution by perception theorists. In expressing thanks, however, I myself must neglect the fact that **TODOROVIČ** (like some other commentators) goes on to express doubts about virtually every particular of the theory I have proposed.

The overview of "basic kinematics" with which **TODOROVIČ** begins is largely consistent, as far as it goes, with the corresponding material developed over a decade earlier in Carlton and Shepard (1990a; 1990b). But TODOROVIČ himself neglects two significant parts of that earlier formulation. The first is our proposal that (under conditions favoring the compelling experience of apparent motion), the motion is experienced over a well-defined *geodesic path* in the abstract space of distinguishable orientations of the object. Yet it is just this proposal that provides the basis for the simplicity and the uniqueness of the spatial transformation experienced, which I claim and which TODOROVIČ questions.

The second part that **TODOROVIČ** neglects is our development (in Carlton & Shepard 1990b) of a characterization of the alteration of the abstract space of distinguishable orientations, and, hence, of the resulting alternative geodesic paths, that is induced by symmetries of the object (or by quantifiable approximations to such symmetries). This latter characterization, which has received empirical support from several experiments (including those of Farrell & Shepard 1981; Shepard & Farrell 1985), together with our proposals concerning the role of local symmetries in determining how the axis of rotation is itself expected to move in certain asymmetric situations, is relevant to other questions raised by Todorovič.

In addition, **TODOROVIČ**'s claim that some of the measured curvatures of paths of apparent motion (including some reported by McBeath & Shepard 1989) fall short of the predictions derived from kinematic geometry, ignores my proviso (mentioned in my response to **BARLOW**) that the conditions must be favorable for effective engagement of the internalized representation of kinematic geometry. Decrease of performance under less favorable conditions does not preclude the existence of a representational competence demonstrated under conditions more favorable to its engagement.

The issues concerning geodesic paths in the manifolds for objects with various symmetries – involving, among other mathematical concepts, the manifold that is the *quotient* of the manifold corresponding to the orthogonal group SO(3) and the manifold corresponding to the object's symmetry group S(O) – are probably too technical to delve into here. (For a full treatment, see Carlton & Shepard 1990b.) For present purposes, I confine myself to indicating a few specific places where **TODOROVIČ**'s characterization of the theory I have advocated is inaccurate or misleading. (The following quotations are all from TODOROVIČ's target article.)

> It can be shown that, given two arbitrary positions of a rigid body, it can always be transported from the first position into the second position by an (almost) unique helical motion . . . . "almost" unique because the same considerations about clockwise and counter-clockwise directions and multiple turns apply. (end of sect. 3.3)

As it stands, this statement is inaccurate. True, the "motion" is not uniquely determined because, as explicitly noted by Carlton and Shepard (1990a), the motion *could* be in either direction and with any number of additional full $2\pi$ turns. (If as we claim, however, the motion tends to be experienced over the shortest path, the motion will tend to be unique except for objects whose orientations differ by an angle close to $\pi$.) The geodesic *path* along which the motion is experienced is nevertheless strictly unique – but only for an asymmetric object. As developed by Carlton and Shepard (1990b), symmetries of the object induce changes in the manifold of distinguishable orientations such that there are alternative connecting geodesics (but, again, with motion tending to be experienced over the shortest paths).

According to **TODOROVIČ**,

> the presumed internalization concerns an invariant or recurrent feature of the environment . . . in Shepard's apparent motion account such a pervasive external regularity is missing. . . . [I]n our world these particular types of [circular or helical] motions do not appear to be typical or representative. Thus there ap-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

> parently is no corresponding pervasive external regularity to be internalized. Consequently, tendencies for perceived circular or helical motions can hardly be based upon internalization of invariant environmental features. (sect. 5.1)

I have two answers to this. The first is the one that **TODOROVIČ** acknowledges in his next paragraph – namely, it is not the "motions" typically occurring in the world that I claim are internalized. Rather, it is the more abstract kinematic geometry of three-dimensional Euclidean space, which provides for the inner construction of the simplest connecting transformation. (But, contrary to TODOROVIČ, in saying this, I do not "concede" anything.) The second answer (already given to BARLOW and to SCHWARTZ) is that circular and helical motions are in fact environmentally prevalent – as local representations of all smooth motions, during sufficiently short intervals of time.

**TODOROVIČ** states that "Shepard's . . . account conveys a portrayal of kinematic geometry and classical physics as two theories that can have different predictions about some aspect of reality similar to, say, Newtonian and Einsteinian theory . . . " and,

> Whereas kinematic geometry describes the ways bodies move, classical physics, accepting this description, goes on to inquire about the physical causes of their motions. . . . For example, kinematic geometry describes the shapes and velocities of trajectories of heavenly bodies, whereas classical physics deduces these shapes and velocities. (sect. 5.4, beginning)

But kinematic geometry does not "describe" or make "predictions about" the "shapes and velocities of trajectories of heavenly bodies." If it were successful in doing this, we would not need physics (or, specifically, the dynamical branch of physics called celestial mechanics) for this purpose. Not being a physical theory and, hence, not including anything about velocities, masses, forces, or accelerations, kinematic geometry neither describes nor predicts actual trajectories of any material bodies. Instead, kinematic geometry is (as its name implies) a branch of geometry, not (as **TODOROVIČ** and some other commentators assert) a branch of physics or (specifically) of mechanics. Kinematic geometry specifies how rigid objects *can* move (or can be moved) in three-dimensional Euclidean space. In addition, it specifies which particular paths of motion are geometrically simplest.

For the purposes of establishing that two different shapes projected on the two-dimensional sensory surface could be of the same three-dimensional object in the world and, especially, for doing so as quickly as possible, kinematic geometry is precisely what is needed, – not information about how particular physical objects typically move in the world. Ultimately, of course, empirical findings should discriminate between alternative proposals (including those proposed in **Parsons**'s commentary). But for this to happen, we need both a full specification of the alternative theories and a careful adherence to the conditions the theories prescribe as appropriate for their test. **TODOROVIČ** does not appear to have proposed or defended an alternative general theory. Moreover, as I have indicated, much of the data that he takes to be disconfirmatory of *my* theory have come from what I regard as inappropriate tests.

### *SR2.4. Hecht*

The article by Heiko **HECHT** is largely devoted to arguing that neither the ways objects actually move, nor the laws of physics governing such motions are accurately represented in the mind. Even if true, this is not directly relevant to my central thesis. As I have emphasized, the phenomena of apparent motion and of mental rotation are to be understood primarily as manifestations of an internalized kinematic geometry, and not of the way objects typically move in the physical world or even of the laws of physics that govern such motions. Nevertheless, there are regularities of the world that derive from physical law, and HECHT is right in suggesting that according to my general argument, to the extent that such motions are both perceptible and biologically relevant, I should expect natural selection to favor their eventual internalization.

My target article began by giving an example of one such physical regularity that has incontrovertibly been internalized. This is the period of the Earth's rotation, which has long held invariant to a (biologically) high degree of precision. Given the high moment of inertia of planet Earth – and, presumably, of any planet capable of supporting the evolution of intelligent life – this invariance is a consequence of the universal physical law of conservation of angular momentum.

In contrast, the principles of general relativity and quantum mechanics – to the extent that they depart from those of Newtonian mechanics – seem not to have been internalized. Indeed, as demonstrated by Proffitt and his coworkers, even the classical Newtonian mechanics of extended body rotation (strikingly exemplified by the behavior of gyroscopes) escapes our intuitive grasp (Proffitt & Gilden 1989; Proffitt et al. 1990). In these cases, presumably, the phenomena have either not been perceptually accessible to our ancestors, and/or have so far not been of sufficient biological significance to have been internalized through natural selection.

From an evolutionary standpoint, in determining what physical principles are internally represented, and to what extents, we must take care to ensure that the experimental tests are appropriate to engage whatever perceptual-cognitive mechanism are likely to have been selected for in our ancestral line. Clearly, paper-and-pencil tests, which presumably were not critical for the propagation of our Pleistocene ancestors, are not ideal for this purpose. As **HECHT** himself observes, humans' "*explicit* knowledge about trajectories of falling objects is seriously flawed" (emphasis mine). (Even paper-and-pencil tasks can, however, engage evolutionarily internalized knowledge of some, perhaps less perceptual, kinds – see, in particular, Cosmides & Tooby 1997.) Also instructive are tasks in which psychophysical judgments indicate distortion of perceived size but the more automatic, lower-level thumb-and-finger grasping response does not (see Milner & Goodale 1998).

I proceed, next, to my observations concerning the specific examples cited by **HECHT**. (For further, insightful consideration and relevant experimental studies concerning these examples, I recommend the commentary by Horst **Krist**.)

If we wish to know how well people have internalized the law of inertia governing the trajectory taken, upon its emergence, by a ball propelled through a semicircular tube, we should be less interested in how people describe or draw the expected trajectory on a sheet of paper than in how people judge the relative naturalness of motions over alternative trajectories that are simulated in the physically most concrete and realistic way possible. We might also be in-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

terested in where a person would be willing to stand relative to the end of the curved tube through which the projectile is about to be fired by compressed air. Similarly, if we want to know how well people have internalized the ballistic trajectory of a dense ball thrown at some upward angle, we should be less interested in how someone describes or draws such a trajectory than in where a baseball fielder actually runs to intercept such a ball.

Likewise, if we wish to know how well people have internalized the principle of statics governing the equilibrium distribution of water in a tipped container, we should not assign great importance to how they rate the naturalness of a line drawing of tipped beakers with lines representing water surfaces drawn at various angles. More relevant would be the outcome of a compelling virtual reality experiment in which people judge the naturalness of a situation in which they themselves slowly tip a beaker they hold in their hand, while the visible surface of the liquid is contrived to appear to depart at different degrees from a strongly instantiated visual and vestibular horizontal.

I would be surprised if the more realistic tests did not generally lead to more physically correct judgments – as was in fact found by Kaiser et al. (1985a) and by Kaiser et al. (1992 – of which **HECHT** himself was a co-author). It may also be that when insufficiently realistic support is provided to engage physical intuition, people revert to other strategies, possibly those using intuitions of the more deeply internalized kinematic geometry.

In the case of the trajectory of the ball fired from the semicircular tube, the physically incorrect circular extrapolation could arise as a manifestation of the acquisition of a "curvilinear impetus" (as suggested by McCloskey et al. 1980). But such a circular extrapolation, as well as the notion of a "curvilinear impetus" itself, could arise because the circular path is the kinematically simplest (as specified by the two-dimensional case of Chasles's theorem, known as Euler's theorem). The apparent motion results long ago reported by Brown and Voth (1937) – which McBeath and I were able to reproduce in the laboratory – are suggestive in this connection. Point-like lights that are successively flashed around the corners of a square against a dark background gives rise, at a certain rate, to the apparent motion of a light whirling around a circle.

Some of the ways in which I deal with evidence that appears inconsistent with my theory leads **HECHT** (and some other commentators) to suggest that my theory may be unfalsifiable. As I indicated in the introductory section, however, a theory is not unfalsifiable if the theory itself makes prescriptions about the conditions under which it is expected to hold. Galileo's theory of falling bodies is not rendered unfalsifiable by the fact that deviations from the law are explainable in terms of interfering circumstances of friction or air resistance. Similarly, the theory that we have a certain geometrical competence is not rendered unfalsifiable by the fact that failures of performance are explainable in terms of such interfering circumstances as perceptually inadequate or unrealistic presentation, excessive demands on memory capacity, or the like.

**HECHT**, like **BARLOW**, **SCHWARTZ**, and some commentators, faults my invocation of kinematic geometry as being "vague," "ephemeral," "imprecise," "unspecific," or "poorly specified." But, the geodesic paths specified by kinematic geometry (or by physical dynamics, or by some interpolation between these) are none of these things. Admittedly, there is room for further clarification of exactly what conditions of test will sufficiently engage the implicit knowledge we do have of kinematic geometry (or physics). What strikes me as vague, is **HECHT**'s Figure 4 – at least in the absence of a clearer specification of what he means by his dimension of "Resolution" or why he deems "Bayesian genericity" wholly unfalsifiable. (The commentary by Hecht's sometime co-worker Mary **Kaiser** also leads me to wonder where in this diagram Hecht would place his own alternative theory of "externalization.")

**HECHT** concludes by saying, of my vision for psychological science, that such "hyper-abstraction leads to immunity [from falsification] and removes internalization from the empirical discourse." But has not a similarly high level of abstraction been necessary for the highest achievements in physics, including general relativity and quantum mechanics? (The "hypotheses with which [physics] starts become steadily more abstract and remote from experience" – Einstein 1949, p. 91.) Hence, what exactly is the difference between the cases of physics and psychology?

### *SR2.5. Kubovy and Epstein*

The wheels of Michael **KUBOVY**'s and William **EPSTEIN**'s analytical engine grind exceedingly fine. I feel I've been left in a cloud of dust. But, if I may stretch *this* metaphor (*pace tua,* Michael!), does the settling dust amount to a substantial mountain – or to a magnificent range of semantical mole hills?

**KUBOVY & EPSTEIN** begin with the "inverse projection" problem – the problem that: *because* different states of the three-dimensional world can project the very same pattern on the two-dimensional receptor surfaces, the actual state of the external world cannot be inferred solely from the instantaneous available sensory information. The question then arises: what is it (beyond the immediate sensory input) that enables the perceptual system to yield up only that one percept which does veridically represent the actual state of the world? **KUBOVY & EPSTEIN** distinguish among perceptual theorists on the basis of their answers to this question as follows:

- For Helmholtz, it is (unconscious) knowledge gained from the individual's previous multi-sensory interactions with the world.
- For the Transactionalists (e.g., Ames, Ittelson, Kilpatrick), it is the (unconscious) "assumptions about the world which assign likelihoods to the candidate solutions."
- For the Constructivists (e.g., Rock), it is the "intelligent" (though still unconscious) use of "internal laws and rules," shaped "over the history of the species."
- For the Computationalists (e.g., Marr, Ullman, Poggio, et al.), it is the "computational modules" that have been "shaped" ("so that their output . . . is adaptive") "over the course of evolution" by the "environmental regularities."
- For James J. Gibson (and the followers of his "ecological" approach), it is the direct "pick-up" of – or (as he has also put it) the "resonance" to – the "invariants" in the "ambient optic array" that converges from the surrounding "spatial layout" onto the eyes of the "freely moving" individual.

**KUBOVY & EPSTEIN** claim to discern major differences among these proposed solutions to the inverse projection problem. Accordingly, in their effort to "locate Shepard's position in this theoretical landscape," they are vexed by



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

my having endorsed, at one time or another, the viewpoints represented even by the extreme ends of this landscape – namely, those of Helmholtzian "unconscious inference" and of Gibsonian "direct perception" (beginning of K&E's sect. 1.2). Yet, the view of perception to which I have arrived can, I believe, accommodate the essential insights of the full range of these proposed solutions. From this tolerant standpoint, the remaining differences among these theoretical positions are, for me, matters more of terminology than of substance; a (relatively) benign neglect of some aspects of perception rather than categorical rejection.

Some may argue that visionary neglects may be so pronounced as to suggest the presence of a theoretical scotoma in the extreme case of Helmholtz or of Gibson. But, is it really plausible – given their awareness both of Darwinian natural selection and of the brain's richness of neuroanatomical structure at birth – that Helmholtz would have categorically rejected the possibility that the perceptual system has been, at least to some degree, evolutionarily shaped; or that Gibson would have categorically maintained that a system can selectively "resonate" to external invariants without having any internal structure to determine its own resonance characteristics?

Granted, Gibson's claim that the information immediately available in the freely moving individual's ambient optic array is sufficient to "specify" the surrounding layout does seem incompatible with the realization, implicit in Helmoltz's principle of unconscious inference, that the mapping of the world to the sensory surface is many-to-one. But, as illustrated in the example I offered in my response to **BARLOW**, the individual veridically perceives the cubical shape of a distant object – even in the absence of relative motion and, hence, even when the optic array does *not* rule out the alternatives of objects with non-cubical shape, curving, non-parallel edges, and non-orthogonal corners! Cases of this kind (however "ecologically invalid" in Gibson's sense) show that perceptual experience is not always fully determined by the immediately available information. Veridical experience depends, in addition, on world knowledge that already exists within the individual – for instance, in this case, knowledge that with probability one, the object is not being viewed from a special angle (cf. Rock's principle of "nonaccidentalness" – Rock 1983).

As for the intermediate positions of the transactionalists, the cognitive constructivists, and the computationalists, I confess that their differences concerning the inverse projection problem strike me as insignificant (despite the pains taken by **KUBOVY & EPSTEIN** to bring them to the fore). The principal difference cited by KUBOVY & EPSTEIN is whether a rule is "followed" (as stated by the constructivists) or "instantiated" (as stated by the computationalists). Certainly, I hold that what I call "internalized principles" are internally instantiated. But, by virtue of their instantiation, they also "actively" affect perception. At least in this sense, they might also be said to be "followed." What I would *not* want to say – though Kubovy & Epstein attribute such an inclination to me (as well as to Rock) – is that these principles are "mental contents." The trouble with terms such as "mental contents" (and also "rule following") is that they suggest that the principles are present to consciousness and are "consciously followed." These are notions that I have expressly denied.

True, I have found merit in Kubovy's (1983) recommendation against language that raises unnecessary ontological issues about the nature of mental entities and operations. In our early reports on "mental rotation" I (and my students) did sometimes use such phrases as "the rotation of a mental image." Such expressions could suggest that we thought that a mental image is the sort of thing that could literally rotate, just as a neurosurgeon's scalpel might physically rotate within a patient's physical head. Anyone who has carefully read what I have written about "second-order isomorphism" or, still more pertinently, about the "analog" character of imagined transformations, however, should recognize that we intended no such thing. Still, we could have minimized the likelihood of such a misunderstanding by using more ontologically neutral language. As Kubovy suggested, instead of saying "The subject rotated a mental image of the object," we might better have said, merely, "The subject imagined the rotation of the object."

As science advances, however, what might at first have been regarded as largely metaphorical can turn out to have a somewhat more literal interpretation. Kubovy's proposed neutral rewording is in fact so neutral that it fails to convey that in the mental rotation studies that Cooper, Metzler, and I reported, there was something actually rigidly rotating. It was not, of course, a concrete physical object, but something very abstract. It was, precisely, the orientation in the physical world in which an appropriate physical test stimulus, if it were presented at a given moment, would lead to a quick and objectively correct decision as to whether that test stimulus was the originally presented object (versus, for example, some slightly altered variant or the object's enantiomorphic mirror image). By "quick" I mean that the overt match-mismatch response was made within about a half-second of the onset of the test stimulus. If the same test stimulus were, instead, presented in any other orientation at that moment, the decision time would increase markedly with the degree of departure from the orientation that would yield the quick response (Cooper 1975; Cooper & Shepard 1973; see especially, Cooper 1976; and, for a more extensive overview, Shepard & Cooper 1982).

What was shown to be literally rotating was not, of course, a literal image (mental or physical). It might be characterized, rather, as the quite abstract, counterfactual conditional possibility of an objective match. It is one of the abstract relational things that, though non-material, actually exist. (A somewhat analogous example would be the relation between two spatially separated but quantum-mechanically "entangled" particles. Prior to performing a measurement operation on either particle, quantum mechanics requires that no spin orientation exists for either particle. At the same time a quantum mechanical conservation principle requires the existence of a conditional *relation* such that, when and if measured, the spin orientations of the two particles – no matter how far separated they may then be in the universe – must be opposite (i.e., if either particle is "spin up," the other must be "spin down," and vice versa).

There may even be something in the subject's physical brain that "rotates" but, here too, it is not a rigid physical object or even a rigid pattern of electrical activity. Again, it is something that is very abstract but in a different sort of way here. Based on electrophysiological recordings from the brains of monkeys, Georgopoulos et al. (1988) reported evidence for the "rotation of a population vector" in the abstract space of possible patterns of neuronal activity. And, in recent priming studies, Kourtzi and Shiffrar (1997;


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

1999a; 1999b) have found positive priming for the recognition of an object presented in a novel position when that position fell within the path of apparent motion but not when it fell outside of that path. So, while I cannot fault **KUBOVY & EPSTEIN**'s Occam-like principle, I suggest that they may have been a bit overzealous in applying this principle to my theoretical accounts of apparent motion and mental rotation.

I would also caution **KUBOVY & EPSTEIN** that the Occam's axe they have to grind is double-edged. Their parting exhortation that "we formulate our theories in as neutral a language as we can," should apply equally to themselves. Yet, in the second paragraph of their article they write: "According to Shepard . . . the fact that apparent movement is *perceived* at all is owing to the 'internalized principle of object conservation' . . . " (my emphasis). I have in fact tried to avoid speaking of "perceiving" apparent motion. For many (including, I believe, most philosophers of perception), the word "perceiving" is taken to mean "veridical perceiving" (a phrase that would accordingly be redundant for those philosophers). For such readers, the apparent motion display may – much as a dream or hallucination might – give rise to the "experience" of motion (as I prefer to say) but not to the "perception" of motion, given that there is (as KUBOVY & EPSTEIN might themselves agree) no physical motion to be perceived.

Yet, I think it would be sad if Kubovy were to give up some of his own wonderful metaphors – including his suggestion, from evidence concerning tactile recognition, that the "mind's eye" is located behind one's own head (Joseph & Kubovy 1994). The evidence Kubovy presents, which is fascinating in itself, is the following: People correctly identify an alphanumeric character traced on the back of their head, even though that patch of skin has had no previous experience in recognizing such characters. However, if the character is traced on the forehead instead, recognition is generally quicker if the tracing is done in mirror image rather than in its normal version. Correspondingly, whereas the letter "b" is immediately recognized as such when traced on the back of the head, it is often identified as the letter "d" when traced on the forehead. If subjects do apprehend the traced character by visualizing it with their "mind's eye," then it would seem that this "mind's eye" is, in effect, located behind their own head. This is an appealing metaphor, and one that even suggests further experimental tests.

There are, however, less fanciful possibilities. Perhaps the subject interprets the identity of a spatial pattern more abstractly, by reference to the left-versus-right organization of surrounding space with respect to the subject's own forward-directed body axis – regardless of whether that pattern is in front of, or behind, the subject. Subjects do sometimes identify the character on the forehead "correctly" – that is, as it was traced by the experimenter. In such cases, the visual-to-tactile transfer might be mediated by prior motor-tactile associations built up from touching, feeling, or scratching one's own head by one's own (writing) hand. A more intriguing and cognitive possibility, consistent with both Kubovy's "mind's eye" metaphor and my left-right-spatial-organization suggestion, is that subjects achieve such "correct" recognitions by imagining a 180-degree rotation in space (in which, for example, the subjects imagine looking back at their own forehead).

Incidentally, I believe, a similar possibility explains why a mirror seems to reverse left and right, but not top and bottom. In fact, it reverses neither of these. What it does reverse is front and back. That is, the mirror shows one's own front seen, so to speak, from the back – as if just the hollow surface of one's front were carried forward, without reversal, to its apparent location behind the mirror. Because the retinal projection of our mirror image is exactly like that produced by a complete, solid human body, and because we have had no experience with viewing hollow fronts of such bodies from the back, we interpret what is before us as the front of a solid human body as seen from the front. But this implies that the body has been rotated through 180 degrees, in order to present us with its frontal view. Because the human body has a left-right symmetry but not a top-bottom symmetry, this implied rotation can only have been around its vertical axis and not its horizontal one. This, I have proposed, is why we experience the illusion that the mirror reverses left and right (e.g., Shepard & Hurwitz 1984, p. 170).

In the end, what I suppose **KUBOVY & EPSTEIN** (along with many other commentators) find most problematic is my notion that universal features of the world – including, particularly, those of kinematic geometry – have been in some way "internalized." KUBOVY & EPSTEIN regard this notion as a mere metaphor, and one that either does no more theoretical work than the equation, "internalize = ingest as food," or else does positive harm (just as Kubovy had maintained that to speak of the "rotation of a mental image" leads people seriously astray).

I believe, however, that in science, as in all aspects of life, our thinking is implicitly guided by metaphors. Of course some metaphors are more apt than others. I venture that the metaphor of internalization, if properly interpreted, may have positive benefits. To **KUBOVY & EPSTEIN** – and to those other commentators who found my notion of internalization problematic – I particularly recommend the commentaries by Margaret **Wilson** and by Gerard **O'Brien** and Jon **Opie**, to which I now turn.

### *SR2.6. M. Wilson*

Margaret **Wilson** has expressed in an admirably clear and sensible way the essential justifications for my use of the term "internalization." If I had done as well in my target article, much of the skepticism about the appropriateness of this term, which runs through so many of the present articles and commentaries, might have been allayed.

I see value, too, in **M. Wilson**'s suggestion that an internal "emulator mechanism" capable of predicting the delayed effects of possible corrections in the control of ongoing behavior may have conferred sufficient benefit to have been internalized. The suggestion is reminiscent of the very early proposal along these lines by Craik (1943), whose historical importance is acknowledged in **BARLOW**'s article. As I suggested in my response to BARLOW, internally instantiated principles of kinematic geometry may play a role in performing such emulations, although the primary role I proposed for such principles is the different one of establishing shape-correspondence most quickly.

### *SR2.7. O'Brien and Opie*

I am grateful, too, for the excellent elucidation of the notion of internalization provided by Gerard **O'Brien** and Jon



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

**Opie**. Clearly, some motions of objects are determined entirely by external constraints that are, in no sense, internally represented within the object. Examples include a planet moving in its elliptical orbit, a toy train chugging around its oval track, and a ball whirling through a circular tube. (Curiously, those people who truly believe that the ball, on emerging from the end of a semicircular tube, will continue in a curved path – as suggested by the experiments of McCloskey, et al. 1980 – must implicitly attribute some internalization of that motion within the ball.)

But, as **O'Brien & Opie** note, there are also cases in which there are no external constraints on the object's path of motion (other than its restriction to some two- or three-dimensional space). In some of these cases the object's motion is actively controlled from moment to moment by its own on-board computer or brain, which may be striving to minimize time, effort, or probability of unfavorable outcome – as in the hypothetical example of the lifeguard seeking to rescue a drowning swimmer, which I presented toward the end of my introductory section SR1. The "functional resemblance" to which **O'Brien & Opie** refer constitutes, I believe, the kind of "appropriate tuning," in that example, of the lifeguard's inner processes to the demands of the rescuing task and justifies the use of the term "internalization." I believe such use justified to the extent that the inner process gives rise to behavior that approximates – even if it does not strictly achieve – the optimum. This brings me to the commentary of **Jacobs et al.**

### *SR2.8. Jacobs, Runeson, and Andersson*

Like these commentators, I have found great value in the insight behind Gibson's ecological approach to perception. Indeed, I regard his insight as being, alone, comparable in significance to Helmholtz's insight about unconscious inference. Moreover, I have resonated to the examples of the "smart" perceptual mechanisms of Runeson (1977), the "smart" behavioral mechanisms of Brooks (1991b), and the "smart" cognitive heuristics of Gigerenzer and Todd (1999). I can appreciate the value of such examples even when put forward (as by Runeson or Brooks) as eliminating the need for "internal representation." Nevertheless, for reasons I have adduced – at least since Shepard (1984), and continuing through my responses to the present commentaries – neither Gibson, Brooks, nor **Jacobs et al.** have convinced me to abandon my conviction that aspects of the world are in some sense internally represented.

According to **Jacobs et al.**, environmental "constraints are taken advantage of by detection of information granted by the constraints rather than by internalizing them." Echoing many of the commentators (and specifically invoking the "ecological approach" and the article by **KUBOVY & EPSTEIN**), **Jacobs et al.** add that they "see no way that such internalized constraints could be beneficial for the perceiver and [hence] no way that evolution could have endowed us with them." Yet, as I tried to illustrate in my response to **O'Brien & Opie** (particularly, in my example of the lifeguard), I maintain the following: (a) "[D]etection of the information granted by the constraints" requires some structure within the perceiver (even if relatively simple – as in the example of the baseball fielder analyzed by McLeod & Dienes 1996). (b) Whatever the neurophysiological implementation of that structure, it must be attuned to those "constraints" and, in this sense, it does internally represent something about them. (c) Because the inner structure thus confers a benefit (such as the ability to minimize time and to optimize outcome), it could indeed be favored by natural selection, as well as by learning.

I heartily agree with the "positive note" of **Jacobs et al.**, that "although individuals might differ in the constraints they exploit, . . . some principles of learning might hold very widely." And when these commentators say, "the minds of individuals are just as likely to reflect local as universal constraints," I can only reply "Of course!" Indeed, "minds" may be *more likely* to reflect local constraints. I am only claiming, first, that perceptual exploitation of constraints (whether local or universal) requires some internal structure appropriate to those constraints (which, in that sense, represents them) and, second, that among the constraints that are thus exploited (and represented), *some* are universal.

### *SR2.9. Heil*

The commentary by John **Heil** provides an especially thoughtful and sensible explication of the contrast between what he describes as, on one hand, "principles on which [internal] mechanisms operate" or on which "an agent's grasp or representation of the principle . . . controls the action" and, on the other hand, "merely . . . principles to which [the mechanisms'] operation apparently conforms" or with which the agent's actions "accord . . . without thereby being . . . guided by that law or principle." This seems to be essentially the contrast that **KUBOVY & EPSTEIN** draw between "following" a rule and "instantiating" it. I would, however, question the advisability of a sharp distinction here. At one extreme, is the hypothetical case in which an agent possesses a complete, perfect, and explicit mental representation of an external situation and computes the theoretically optimum behavior for dealing with it (perhaps even by means of "mechanisms for solving differential equations in the brain," to use **Heil**'s words). At the other extreme, is the case in which the behavioral path taken by the agent is wholly determined by external constraints (like a person strapped to the careening seat in an amusement park ride).

Cases of interest to cognitive/behavioral science typically fall somewhere between these extremes. In the hypothetical example of the lifeguard trying to reach a drowning swimmer (given above and in sect. SR1), there certainly are physical and, in a sense, "external" constraints. Lacking wings, the lifeguard is essentially confined to a two-dimensional surface. If she had wings like a bird, her quickest route to the swimmer would be along the straight connecting line, "as the crow flies." Or, if she had a body adapted, like the seal's, to the aquatic medium, she would be able to move faster through the water than across the sand. Her least-time path would then be one that at the water's edge, deviates from a straight line in the opposite direction from the path that is humanly quickest.

Notice, however, that for all three of these hypothetical cases – that of the bird, the seal, and the human – the three corresponding least-time paths (straight, left-bending, and right-bending) are determined alike by Fermat's least-time principle and yield alike a path in compliance with Snell's law of refraction. Given that the direction in which the lifeguard sets off across the sand is in no way physically constrained, that direction must, I argue, be determined inter-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

nally. But this leaves open questions about the conscious or unconscious, analog or symbolic nature of the internal process, and about the extent to which it accurately "mirrors" or only approximates the whole situation and its optimum solution. (I return to these questions in my reply to **Todd & Gigerenzer** at the end of sect. SR5.) Further conceptual clarification may be obtained from other cases that are intermediate between the two extreme cases mentioned above, differing in the relative amounts of internal (versus external) control, in the nature of the computational process, and in the degree of optimality of the result.

### *SR2.10. Krist*

I welcome Horst **Krist**'s developmental perspective and his citation of empirical studies that support the notion of internalization – particularly vis-à-vis the criticisms leveled by HECHT. I have only minor comments on the following statements in his commentary.

**Krist** writes: "Internalization and modularization are by no means mutually exclusive . . . [though] HECHT appears to take for granted that internalized principles can be revealed in perception, action, imagery, and problem-solving tasks alike whenever the situation is somehow underspecified." In agreement with this, I have repeatedly emphasized that the internalized knowledge underlying perception may be largely inaccessible to introspection and explicit reasoning (and, hence, to paper-and-pencil assessment). At the same time, however, I have conjectured that the mechanisms that may initially have evolved in the service of immediate perception may later have been recruited and elaborated in the service of mental capabilities that are increasingly independent of immediate sensory support. Examples include the capabilities: (a) for experiencing rigidity-preserving (long-range) apparent motion; (b) for imagining spatial transformations of objects that may not even be present; (c) for problem solving; and (d) for the use of thought experiments in arriving at explicitly formulated principles (discussed in sect. SR6). At a given stage of phylogenetic or ontogenetic development, an individual may have some of the more concretely perceptual-like capabilities but not yet the later-emerging, "higher" cognitive capabilities.

**Krist** states: "[K]inematic geometry does not prescribe any particular trajectory for objects exiting curved tubes, and there is no simplest path in this case." As I conjectured in my replies to HECHT and to **O'Brien & Opie**, however, the reported erroneous tendency of many people to extrapolate the trajectory of the exiting objects as a continuation of the curve may in part reflect the kinematic simplicity of circular motion.

### *SR2.11. Kaiser*

I share Mary **Kaiser**'s reservations about HECHT's alternative theory of "externalization." Her statement that "the logical opposite of internalization (as well as externalization) is an unconstrained perceptual system – one that imposes no assumptions, and finds any under-specified stimuli ambiguous and uninterpretable," forcefully expresses what I have been trying to say in response to the more radical followers of Gibson's ecological approach who sometimes speak as if a perceptual system need have no inner structure.

But, I offer two minor qualifications to **Kaiser**'s statement that "Gibson . . . reminded us that Euclidean geometry is not an appropriate description of our visual environment – rather, our world is filled with meaningful surfaces." First, Gibson would surely allow that we experience (in addition to the surfaces of objects) the solidity, heft, and a variety of other "affordances" of the objects. Second, Euclidean geometry may be an appropriate description of what (in a Kantian vein) we might say is *presupposed* by a system that effectively represents the possible motions of objects that three-dimensional Euclidean space affords.

### *SR2.12. Parsons*

I approve of Lawrence **Parsons**' recommendation that alternative candidates for paths of apparent motion and imagined transformations be compared with respect to explicit criteria of "efficiency and utility." In his commentary, Parsons considers what I take to be the following three types of trajectories between any two different spatial positions of a given object (which I shall assume to be globally asymmetric, to ensure uniqueness of kinematically specified axes of rotation – see Carlton & Shepard 1990b):

1. *Screw displacement* – minimum rotation together with a concurrent translation on the same fixed axis, which is uniquely determined (for the two object positions) by kinematic geometry alone, that carries the object from the one position to the other.

2. *Shortest trajectory* – minimum rotation around an axis (whose orientation, only, is uniquely determined by kinematic geometry), together with the concurrent rectilinear translation of that axis that carries the object between the two positions.

3. *Dynamical spin-precession* – rotation around one of the object's own principal axes of inertia, which axis concurrently rotates around a second axis that is oriented in an environmentally salient (e.g., vertical or horizontal) direction, and that also concurrently translates so as to carry the object between the two positions.

Naturally, these different types of trajectories will differ with respect to particular criteria of "efficiency or utility." To the extent that the proposed criteria capture something that has been of biological significance for our ancestors or ourselves, our perceptual-cognitive representational system may have come to perform (through natural selection and/or learning) the corresponding type of transformation. Such an outcome would be wholly consonant with my general approach. Indeed, this is why, in Carlton and Shepard (1990b), we explicitly developed a one-parameter family of models varying continuously between the first two above-listed cases – that of the purely kinematic "screw displacement" and **Parsons**' more physical-dynamical "shortest trajectory." It is also why we formulated, as Parsons did for the third type listed above, a type of trajectory that can be anchored to salient features of the environmental frame. But, for this latter purpose, we proposed, instead of Parsons' more dynamical "spin-precession" alternative, the following purely geometrically defined type of trajectory:

4. *Kinematical spin-precession* – rotation around one of the object's own *axes of local symmetry* which concurrently rotates around a second axis that is oriented in a perceptually salient (e.g., environmental) direction and that *also* concurrently *translates* so as to carry the object between the two positions.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Our version of "spin-precession" has the following possible advantages: It is defined purely in terms of visually available geometrical properties, whereas **Parsons**' dynamical version invokes the physically defined principal axes of inertia, which are not directly given by information available at sensory surfaces (or in the ambient optic array). Such physical properties of objects depend on assumptions about the density distribution internal to the object, and remain undefined for many familiar objects (even assuming uniform density). (For instance, the cube has no principal axes of inertia defined.) Moreover, there are empirical indications that the rotational dynamics of extended bodies (such as gyroscopes which dramatically manifest a dynamical "spin precession") is poorly represented internally, if at all (Proffitt & Gilden 1989; Proffitt et al. 1990; Shiffrar & Shepard 1991). As illustrated in Figure 2 of my target article, the axes of local symmetry are perceptually salient and intuitive in a way that principal axes of inertia are not.

### *SR2.13. Pani*

Like Parsons, John **Pani** points to the importance of "the salient reference system" for determining which spatial transformations are likely to be perceptually simplest or mentally performed. The issue that this valid concern principally raises for the type of theory I have been trying to develop is this: What is the most natural way in which the existence of environmentally salient directions might be incorporated into the geometry of the abstract manifold representing orientations and rotations?

So far, as noted in my preceding response to **Parsons**, I have explicitly acknowledged the role of external reference frames in perception and cognition – including, especially, that of the invariant gravitationally conferred vertical (see, e.g., Shepard 1982b; Shepard & Hurwitz 1984; Shiffrar & Shepard 1991). Though not cited by **Pani**, the Shiffrar and Shepard study in particular demonstrates a strong and highly orderly dependence of the perceptual experience – and the accuracy of recognition – of a cube's rotation on the orientation of the rotational axis relative to the environmental vertical and horizontal. It also demonstrates a comparably strong and orderly dependence on the orientation of the rotational axis relative to the symmetry axes of the cube. (Fig. 2 of my target article is confined, however, to illustrating only the latter, symmetry effect, for which a theoretical basis is more fully provided in Carlton & Shepard 1990b.)

### *SR2.14. Intraub*

The spatial layout of the surrounding environment provides, of course, an important reference frame. Moreover, this layout provides the "local constraints" that, first and foremost, are, according to Gibson, "picked up" from the ambient optic array. Helene **Intraub**'s commentary calls to mind a striking thing about this layout. We demonstrably have knowledge of large portions of the layout and its "affordances" that are not at the time represented on our sensory surfaces. For example, with attention focused on the work before me, I reach back without turning and successfully grasp a book or a bottle of water I had recently placed on a table top to my side; or, to check the time, I turn right around and look immediately at the location behind me where a clock has long hung on the wall. I find it natural to say that these and other adaptive behaviors are guided by information that, though not currently available in the sensory input, is represented internally.

I do not, however, believe that my "theoretical framework" is as "susceptible," as **Intraub** seems to fear, to the "concerns" raised by **TODOROVIČ**, **SCHWARTZ**, and **HECHT** (including SCHWARTZ's worry about ecological validity). But I value her demonstrations that in the memory representation of a scene previously viewed in a photograph, objects that were partially cut off by the edges of the photograph are unconsciously completed and the background accordingly extended. In addition, I embrace her conclusion that this normally occurring "error" provides evidence for internalized knowledge; and I second her answer to HECHT's question about falsifiability – namely, that it is possible to "articulate a 'boundary condition' for this boundary extension."

I have long believed that the evolved capacity for remembering spatial layouts in humans (and to impressive degrees in many other species – including such birds as the nuthatch), provides a mental framework for orienting ourselves and keeping track of things that are not always visible in our environment. The breakdown of this ability that can result from certain types of brain damage precipitates what was has been called (by Goldstein, as I recall) a "catastrophic" reaction. I've experienced something of the feeling myself when workers who (having sought access to wiring or plumbing in the ceiling of my office) have haphazardly moved many stacks of my papers and books.

I have also been struck by how people normally make use of spatial locations as mnemonic place holders. In referring to a concept that had been introduced earlier in a meeting, they quite unconsciously gesture toward the location on the blackboard where the concept had been diagrammed, even after the diagram has been erased; or they gesture toward the place around the seminar table where the person who had introduced the concept had sat, even after that person has left the room.

### *SR2.15. Vallortigara and Tommasi*

Another example of representational completion is the visual phenomenon that Giorgio **Vallortigara** and Luca **Tommasi** present in support of a proposed universal perceptual rule for deciding which of two objects that project overlapping retinal images of indistinguishable color is in front of the other in the three-dimensional world. These commentators argue that this rule (which they attribute to Petter) reflects "the geometrical property that when in overlapping objects larger surfaces are closer, there will be shorter occluding boundaries than when smaller surfaces are closer." Internalization of this rule explains why, in their illustrative figure, the black body of the hen is generally seen to be in front of the black fence, even though the T-junctions where the fence crosses the hen's (white) legs causes the legs to be perceived as definitely behind the fence.

These commentators indicate that evidence has been obtained (by Forkman and Vallortigara) for the operation of this same perceptual principle in birds, whose visual system is anatomically quite different. As I mentioned in section SR1.1, this may be an example illustrating convergent evolution of a functional match to the world, despite differences in gross anatomical structure.

**Vallortigara & Tommasi** imply that people are some-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

times briefly able to overcome the occluding-boundary-minimizing principle (perhaps when striving for consistency of depth interpretation while focusing primarily on the legs of the depicted hen). But they don't mention what strikes me as a quite remarkable illusion of object completion. Perhaps I have an exceptionally vivid imagination, but when I was able, during brief periods, to see the body of the hen as located behind the fence, I had the unmistakable visual experience of the fence continuing in front of the hen, even though both the fence and the hen are objectively of the same uniform black color.

### SR2.16. A. Wilson and Bingham

The commentary by Andrew **Wilson** and Geoffrey **Bingham** raises, again, the issue brought to the fore by **HECHT** and by **Parsons**, of the relative roles of kinematic geometry versus physical dynamics. I am, in fact, in complete agreement with **Wilson & Bingham** that we perceive (or "pick up") the "affordances" of many objects from the dynamics of their interactions in the world – rather than from any static snapshot. Of course we recognize a dropping spherical object as a rubber ball from the way it bounces off the floor, as a hollow metal shell from the way it dents without bouncing, as a solid ball of heavy, rigid material from the crater it makes in the sand, and as a viscous glob of doughy material from the way it flattens on impact into a pancake shape. But my target article was not concerned with the perception of such affordances of objects. It was concerned, rather, with the kinds of rigid motions that are afforded by three-dimensional Euclidean space and, also, with the quickest transformations that afford comparison of the shapes of differently oriented objects.

**Wilson & Bingham**'s pronouncement that the theory I propose "fails to successfully capture the essence of the perceptual tasks [Shepard] expects of it, such as object recognition," in presuming to have an adequate knowledge of what Shepard "expects," misses its mark. Surely, the whole implication of their commentary – that the existence of one kind of perceptual-cognitive capability excludes the existence of another – is unjustified. Also unwarranted are attributions such as the following: "Shepard uses his findings from studies on thinking to make claims about the nature of perception." (In fact, I have come to see an important, analogical type of thinking as having evolved from mechanisms that had previously evolved in the service of perception.) "Shepard [claims] that only kinematics, and not dynamics, is visually specified and therefore available for internalization" and suggests "that forces in nature are arbitrary." In quoting my statements out of context, **Wilson & Bingham** attribute to me over-generalized, categorical claims that do not well represent the views that I in fact hold.

### SR2.17. Bertamini

Although finding the experience of apparent motion over a curved path to be "fascinating," Marco **Bertamini** aligns himself with the skepticism about internalization of kinematic geometry expressed by **BARLOW**, **TODOROVIČ**, **HECHT**, **KUBOVY & EPSTEIN**. **Bertamini** asserts that the only possible evidence for internalization of kinematic geometry is simplicity, which he takes TODOROVIČ as having shown to be dependent on the (by implication) arbitrary way in which "the problem is formalized." In contrast, I suggest that the results of Bertamini's interesting experiments are essentially explainable in terms of the non-arbitrary symmetry principles described by Shepard (1981b; 1984), experimentally demonstrated by Farrell and Shepard (1981) and Shepard and Farrell (1985), and developed in mathematical detail by Carlton and Shepard (1990b).

### SR2.18. Foster

I am of course in complete sympathy with the fundamental role that David **Foster** assigns to group theory. Indeed, I was in part inspired by the early representation by Foster (1975b) of apparent motion as "minimum-energy" paths in the rotation group SO(3). More generally, I agree with Foster that successful analysis "depends critically on choosing appropriate perceptual representations . . . based on the natural group structures of the spaces involved." I, too, am hopeful that this general approach can be usefully extended to the representation of plastic and other nonrigid transformations of shapes, revealing "a close relationship between apparent motion and visual shape recognition" (see, e.g., Carlton & Shepard 1990b; Shepard 1981b; Shepard & Farrell 1985; and for the related idea that the shape of an object may be represented in terms of a history of transformations from a canonical, simpler shape, see Leyton 1992).

I also share **Foster**'s optimism about applying such an approach to the representation of objects and their common principles of transformations in seemingly disparate domains. The Gestalt grouping principles of proximity and of common fate, for example, operate just as powerfully and in exactly the same way in auditory pitch as in visual space (e.g., see Jones 1976; Shepard 1981a; 1999). Lakatos and I have demonstrated that the time-distance law for apparent motion in visual space also holds within auditory and tactile spaces (Lakatos & Shepard 1997a; 1997b) and, in as yet unpublished work, even in the more metaphorical "spaces" of color (specifically around the hue circle) and auditory pitch (specifically around the "chroma" circle). I return to Foster's own proposal concerning the application of group theory to color in section SR3.

### SR2.19. Hoffman

William **Hoffman**'s applications of group theory to visual perception, including apparent motion, (e.g., Hoffman 1978) also provided an early indication of the potential specifically of Lie groups, which are central to the theoretical formulation later developed by Carlton and Shepard (1990a; 1990b). Unfortunately, those of us who believe group theory to be fundamental to understanding perception and cognition have not done all that we might to make the nature and relevance of such abstract mathematical structures – previously little applied in the behavioral, brain, or cognitive sciences – clear and compelling to less mathematically oriented researchers in these fields. I myself might benefit from a better understanding of Hoffman's proposals concerning the applications of group theory to the representation of motion, as well as to color constancy (the topic to which I return in sect. SR3).

### SR2.20. Frank, Daffertshofer, and Beek

I have not yet studied the self-organizing-process theory advocated by **Frank et al.** in sufficient depth to respond



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

to their commentary with any confidence. I can only offer a couple of possibly superficial observations. First, much as I said in my lengthy response to **BARLOW**, it is not clear to me how such things as "the interaction of inhibitory and excitatory neurons of the visual system under particular boundary conditions" gives rise to the effective imagining of extensive, shape-preserving spatial transformations in the absence of visual stimulation (including, by the way, such transformations as have been tested through other modalities, including touch). Second, the prediction of "a significantly increasing variance of the quality index of apparent motion close to critical SOAs" (though generally observed in work I have done with several co-workers, including Farrell) is presumably to be expected on any plausible theory.

### *SR2.21. Lacquaniti and Zago*

Investigations into what may be internalized about physical – as opposed to purely geometrical – aspects of the world are certainly desirable. To study what is implicitly known about gravitationally accelerated bodies by collecting behavioral and neuroelectric data from individuals who endeavor to catch falling objects, as in the approach described by Francesco **Lacquaniti** and Mirka **Zago**, is a good example of using a test that (unlike a paper-and-pencil test) is likely to engage whatever representations may have been internalized during evolution or individual learning in the world. In addition, the extension of such research to conditions of micro-gravity never experienced by our ancestors (as in the work they cite by McIntyre et al. 1999) illustrates how conditions that are, in this sense, not "ecologically valid," can be of value in establishing whether constraints that have always prevailed in our environment are indeed internally represented in the absence of those constraints.

### *SR2.22. Hood*

As should by now be clear, I accept neither the premise evidently adopted by Bruce **Hood** (following **HECHT**) that phylogenetic internalization is virtually "unfalsifiable," nor the premise that ontogenetic internalization is "anathema" to me. Yet, for some of the reasons I just mentioned in my reply to **Lacquaniti & Zago**, I very much approve of investigations into the emergence of implicit knowledge through developmental studies, such as Hood reviews, in which young children are tested with actually falling bodies.

Of course, we must always be vigilant in interpreting the failure to obtain evidence for the genetic transmission of any particular world knowledge. It could be that that knowledge is in no way genetically transmitted and so, can only be gained through learning. But it could also be that some part of the knowledge *is* genetically transmitted but is expressed only after some level of maturation is attained and/or when activated through relevant experience. Hence, even if it can only be learned, the learning itself (as I have noted) is necessarily guided by principles that must have been genetically transmitted.

### *SR2.23. Niall*

Keith **Niall** begins by comparing (rather too grandly) my thinking that I "had found the kinematics of mind" to Frege's thinking that "he had reduced arithmetic to logic." **Niall** then proceeds to suggest that he himself can reduce my "kinematics of mind" to merely "the characteristics of illumination, or the perspective geometry of pictures." Niall's Figures 3 and 4 do seem to show an approximately linear trend when his quantity – "1.0 – correlation of gray levels" in two-dimensional pictures of a side-illuminated Shepard-Metzler object – is plotted against angular difference in the orientation of the portrayed object. Closer inspection reveals, however, that the impression of linearity comes more from the straight line he has drawn through the plotted points than from the points themselves. For those points clearly follow an inflected curve that systematically departs from the linear function much more than the residual, unsystematic departures in reaction-time data of Shepard and Metzler (1971) or of Cooper (1975; 1976). While admitting that his own "very simple correlation . . . is not likely to account for results on all the . . . experiments on mental rotation," Niall nevertheless takes it to be indicative of the broader possibility that "it is the similarities of pictures, and not the kinematics of representations, that are key to understanding the mental rotation effect."

I shall be curious to see how **Niall** does propose to "account for results on all the . . . experiments on mental rotation." Among others, these include not only the original results of Shepard and Metzler (1971), in which there was no shading. More crucially, they include: (a) the subsequent results reported by Metzler and Shepard (1974), in which critical time was shown to depend not on the relation between the two stimuli compared, per se, but (as in subsequent experiments on apparent motion as well) on which of two alternative paths of transformation between them the subject mentally traversed; (b) the related results of Kourtzi and Shiffrar (1997; 1999b), in which recognition was found to be primed only for test stimuli corresponding to orientations along the path traversed; (c) the results of Cooper (1976), in which correct responses were quickly made only when the test stimulus was presented in the orientation at which, according to that subject's previously measured rate of "mental rotation," the subject should be imagining the object at the unpredictable moment of the test; and (d) results like those reported by Georgopoulos et al. (1988), that the neurophysiological process passes through states corresponding to intermediate orientations.

### *SR2.24. Vickers*

In concluding this section concerning the representation of spatial transformations, I want to acknowledge the potentially enormous power and scope that I see in the "generative transformational approach to visual perception" sketched by Douglas **Vickers**. He is, I think, too modest in suggesting that that approach "can provide a computational model, in terms of which Shepard's internalization hypothesis *retains* its generality" (my emphasis). He might justifiably have said, rather, "*gains* its generality"! For, where I have focused on the establishment of the identity of a single object through generating a rigid spatial transformation from some canonical or comparison object, Vickers has indicated how a whole complex, natural scene might be economically encoded and internally represented through a nonlinear concatenation of such elementary spatial transformations. I


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

look forward, with great anticipation, to the full, fractal "flowering" of such an approach to perception.

## SR3. On the representation of surface colors

### *SR3.1. Brill*

Michael **Brill** is of course entirely correct that "people report scene colors differently if asked 'what is the color of the light?' as opposed to 'what color is that surface?'." The perceptual system does indeed have "multiple levels," and the different kinds of information represented at these different levels can sometimes be tapped by different kinds of questions. (In this respect, the perceptual system is like the higher-level cognitive system, which also has multiple levels of representation, with the result that people give very different answers when asked, for example, to judge the similarities of numbers with respect to visual appearance, sound, numerical magnitude, or abstract arithmetic properties – see Shepard et al. 1975.)

Generally speaking, it is the invariants in the external three-dimensional world that have been of the most significance for our ancestors. This is why our perceptual systems are attuned to the constancies of size, shape, and color of external objects – leaving us less aware of the absolute shapes, sizes, and spectral compositions of the retinal projections of the external objects. Thanks to shape constancy, for example, we perceive the top of a table as rectangular, and can detect relatively small deviations from rectangularity, even though projections of the table top on our retinal surfaces are generally quadrilaterals with very non-parallel edges and very non-orthogonal corners that vary enormously with our position relative to the table.

Yet, these varying aspects of the retinal image are not wholly irrelevant for our behaving in the world. Taking a different attitude, we can also "see" that the far end of the long table "looks" (in a different sense) smaller than the near end. Why would we have evolved the capability of "seeing" in this way if the retinal projection is irrelevant to how we behave in the three-dimensional world? The answer is that the retinal image is highly correlated with some behaviors in the world and, to that extent, is relevant. If I attempt to point to one and then the other corner of the table at the near end, I find I must move my arm through a larger angle than if I attempt to point to one and then the other of the corners at the far end. (For more on perception of table tops, see Shepard 1990b, p. 48.)

Similarly, although it is important to recognize an object as being the same object despite the very different spectral compositions of the light it reflects to our eyes under different conditions of natural illumination, there can be circumstances in which it is relevant to judge the conditions of illumination and even how the "color appearance" of a surface is affected by that illumination. So in the case of color, too, we have gained the capabilities of perceiving, separately, the inherent characteristics of an object and the circumstances under which it is viewed. We see, for example, that the face of a loved one glows redder because it is illuminated by the light of the setting sun or from the fire in the fireplace and not because the person herself is flushed with exertion, fever, or embarrassment.

As alternatives to the linear basis-function model I had adopted from Maloney and Wandell (1986), **Brill** proposes two models for invariant color representation based on quite different assumptions (such as, that reflectance spectra are "Gaussian in a monotonic function of wavelength" or, that the spectral sensitivities of the sensors approach "delta functions in wavelength" and that the illuminant spectrum has a certain exponential form). For many vision researchers, the principal question about the assumptions of such an engineering approach is likely to be: How well do they describe the actual visual mechanisms of humans or other animals? For me, the more fundamental question is: How well do such assumptions characterize the problem faced by any color-representing being or agent in the terrestrial environment (or, more generally, in the environment of any planet conducive to the evolution of highly developed forms of color vision)?

I claim that for any model (or mechanism) yielding fully color-constant representations in typical natural environments of this kind, the input should have at least three chromatic degrees of freedom of sensitivity at each location on its sensory surface. In addition, I claim that although the emergence of color representations of higher dimensionality might be favored under particular circumstances (such as a need to discriminate among particular edible versus poisonous plants, or – through runaway sexual selection – the need for displaying and perceiving elaborate bodily colorations), three dimensions will generally suffice to yield a good approximation to color (and/or lightness) constancy.

### *SR3.2. Gold*

I agree with Ian **Gold**'s conclusion that even if I am "correct in positing internalized principles that facilitate the perception of color, at least some of these principles are likely to be specific to particular species and niches rather than uniform across all animals that perceive color." Color constancy, in particular, may or may not be the most important benefit of color vision for a particular species – such as one whose survival primarily depends on its ability to detect red fruit against a background of green leaves (a case I explicitly considered in Shepard 1994, reproduced in my target article here). And sensitivity to color may have arisen before the achievement of color constancy.

Nevertheless, for veridical representation of surfaces in the external world, there is a sense in which perceptual constancy may be more fundamental than color representation, per se. I have used the hypothetical example of an individual (whether an animal or a robot) for whom color, as such, confers no benefit, and for whom a perceptual representation of surfaces solely in terms of lightness levels (as in a "black-and-white" photograph) is wholly adequate. The visual system must nevertheless analyze its input into three chromatic channels in order to achieve a final representation (even a merely "shades-of-grey" one) that attains lightness constancy and, hence, facilitates object recognition under different conditions of natural lighting.

Granted, a high degree of constancy (whether of color or merely of lightness) may not be essential for some animals. Some species may manage adequately with fewer than three chromatically distinct classes of color receptors. (Indeed, many animals survive and reproduce without any visual receptors at all – though not, generally, macroscopic animals that move about under the natural conditions of illumination prevailing above the surface of the Earth.)



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

### SR3.3. Bruno & Westland

I also agree with Nicola **Bruno** and Stephen **Westland**, that "a complete characterization of the challenges faced by color perception must include [other complications arising in 'cluttered environments,' including] . . . illuminant changes due to inter-reflections between surfaces . . . ." As I indicated in the introduction (citing Galileo's discovery of physical laws by abstracting away from such complicating factors as friction, air resistance, gusts of wind, etc.), I have been striving for the most simple and general principles that emerge in the absence of complicating circumstances. I suspect that such factors as "inter-reflections between surfaces," though certainly demonstrable and interesting, are of secondary importance even in most natural environments. (Nor is it clear that such inter-reflections cannot be corrected for within a three-dimensional compensating framework of the general type I have advocated here.)

In any case, the striking fact remains that under natural variations of illumination, we do achieve generally good color constancy with just three opponent dimensions of color representation. Moreover, as I indicated in my preceding response to **Gold**, although it may be that color vision in many vertebrate lines evolved (as suggested in my target article as well as by Gold and by **Bruno & Westland**) "to allow the detection of brownish-red edibles against greenish backgrounds," color constancy is more fundamental than color registration per se in that even achromatic lightness constancy requires that the input be analyzed into three chromatic channels.

### SR3.4. Sokolov

Given my long-standing quest for appropriate ways in which to represent psychologically significant objects and qualities as points in a representational space, I am of course quite sympathetic to attempts to formulate spatial models for colors in particular. The proposal of E. N. **Sokolov** (and his collaborators, especially C. A. Izmailov) for representing colors on a three-dimensional spherical "surface" embedded in four-dimensional Euclidean space is accordingly of considerable interest to me. In the present context, however, I have a few reservations concerning Sokolov's particular proposal.

First, and most fundamentally, **Sokolov**'s focus on the neuronal level of color representation, leaves open what is for me the central question: What in the world (if anything) can be identified as rendering a proposed representation (and its neurophysiological implementation) particularly adaptive for the organism? I have argued that a three-dimensional opponent-process representation – which, incidentally, entails the circular representation of hue – may have been selected principally for the following reason: It allows for correction of the naturally occurring variations in illumination and, hence, for the achievement of the color constancy and, thence, for the recognition of significant objects in the world. This is not inconsistent with the representation proposed by Sokolov, which is also an opponent-process representation with an intrinsic dimensionality of three. But I would like to see a theoretical analysis of what it is about the external world (and not just about the internal neuronal network), that confers an advantage on organisms that represent colors in a three-dimensional space having a pronounced intrinsic positive curvature.

**Sokolov** and Izmailov have reported empirical evidence for such a hyperspherical representation by applying multidimensional scaling (MDS) *both* to neurophysiological data and to human subjective judgments of color similarities. This is intriguing to me – in part because my associates and I have argued that the positions of objects in space also correspond to points on the three-dimensional "surface" of a hypersphere (see my target article here, and Carlton & Shepard 1990a). At the same time, I realize that I am not entirely clear what answers Sokolov would give to the following two questions: Why have applications of MDS to judgments of color similarity reported by several other vision researchers not revealed a gross inadequacy of an essentially flat, Euclidean three-dimensional representation of colors? And why do subjective color differences correspond, in his representation, to direct distances through the four-dimensional embedding space rather than to the geodesic distance within the three-dimensional space itself?

Finally, **Sokolov**, – in mentioning his finding (with Izmailov) that discriminative reaction-time falls off with "subjective color differences" (as determined by MDS) according to a reciprocal or hyperbolic function, – says that the form of this function is explained by the way in which color differences are "computed" in the neuronal network. But, in my theory of generalization (Shepard 1987b), I argued that a reaction-time function of just such a form is to be expected for discrimination of stimuli of any sort. I wonder about the need of a theory that is specific to the representation or computation of differences in color.

### SR3.5. Foster

The generalized group-theoretic approach that David **Foster** has taken to the problem of representing object motion evidently extends, as well, to the type of "relational color constancy" that he proposes. This relational constancy appears to be in the spirit of the psychophysical "relation theory" that I (and David Krantz) have advocated (see Shepard 1981c; and, for something about the history of this idea, Krantz 1983). (It is also related to what I have called "second-order isomorphism" – Shepard & Chipman 1970; Shepard et al. 1975.) In the absence of other information, however, it is not clear how a purely relational color constancy such as the one Foster has proposed would yield a specification of the absolute color of any surface in a scene (Laurence Maloney, personal communication of May 28, 2001).

In any case, **Foster** concludes that the "parallels between these various perceptual domains may not be consequences . . . of adaptation to specific properties of the world . . . [as much as to] common organizational rules." From my standpoint, however, this raises the question of what, if any, may be the ultimate, non-arbitrary source of such "common organizational rules." My own tentative answer is that the "properties of the world" to which I have referred should be regarded as including the abstract relations of mathematics just as much as (or perhaps even more than) those of physics.

### SR3.6. Hoffman

Regarding the possible application of group theory to color representation, my curiosity has also been piqued by the seeming formal parallelism between William **Hoffman**'s proposal concerning color constancy in terms of "the quo-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tient group SO(3)/S(O) over the Newton color cone," and the application that my associates and I have made of SO(3)/S(O) for representing the manifold of possible positions and rigid transformations of symmetrical objects in three-dimensional space (see Carlton & Shepard 1990b, and my present target article).

### *SR3.7. Decock & van Brakel*

Philosophers Lieven **Decock** and Jaap **van Brakel** paint a very bleak picture of the prospects for formulating general laws of color representation. Their pessimism seems to stem in part from an unwillingness to take the step that (as I note once again) proved so effective in the discovery of general laws in physics. This is the step of initially focusing on very pure, simple, and well-defined situations, and ignoring the various deviations that (even if small) undeniably occur under less constrained conditions.

Additional problems arise for these commentators because they neglect to distinguish between the psychological, on one hand, and the physical (or the physiological), on the other. Thus they repeatedly conflate two fundamentally different kinds of representational spaces: (a) spaces of the physical stimuli or external objects defined solely in terms of physical measurements on such stimuli or objects, and (b) spaces of the internal representations of such stimuli or objects defined solely in terms of responses of experiencing subjects. Psychophysicists need both kinds of spaces to discover the mapping between them (Shepard 1981c). But physicists can discover general laws without any recourse to the second, psychological kind of space; and psychologists can discover general laws without any recourse to the first, physical kind of space. Indeed, one of the advantages for psychology of multidimensional scaling (Shepard 1962b; 1980) is that it can yield precise results without depending (as in the case of color) on any "precise measurements by means of spectrometers [or] underlying physical theory."

**Decock & van Brakel**, however, imply that for me, "the distinction between phenomenal, perceptual, psychological or internalized representational colour spaces and the various technological or (psycho)physical colour spaces is blurred," and that I take "as self-evident that these colour spaces are isomorphic." This is not an accurate characterization of my view. I regard the relations between any such spaces as a matter for empirical investigation. For this purpose, an explicit specification of the operations used to construct each space is essential. Decock & van Brakel do not themselves distinguish between what they term "phenomenal, perceptual, psychological or internalized representational colour spaces."

As implied in my response to **Brill**, I recognize distinct differences between (a) judging intrinsic surface color, (b) judging the different color appearances of surfaces that are perceptually attributed to conditions of lighting, and also (c) judging the color of light itself – as it arises from sources, specular reflections, and the like. The section on color in my target article focused, however, on the representation of surface color, which I regard as of primary importance. In any event, the data in each of these three cases, being derived entirely from human judgments and responses, provide evidence about the structure of a corresponding internal representational space (whether it is termed "phenomenal," "perceptual," or "psychological"). But all three of these cases are to be sharply distinguished from any "physical" space that is based, instead, on measurements by means of physical devices such as spectrometers. And most certainly, "the internalized color space" is *not* to be "simply equated with a wavelength mixture space." Of course, psychophysical research seeks to discover the functional form of the transformation that converts such a physical space into something that approximates one of the psychological spaces. But the ultimate criterion of the psychological usefulness of such a transformed physical space must be its match to the independently obtained psychological space.

Finally, a number of **Decock & van Brakel**'s specific statements stand in need of correction. Here, I confine myself to this one passage: "[E]ven if MDS techniques yield three dimensions, there is nothing to tell you how to define the axes and measure distances. [Moreover] it has been claimed that four, six, or seven dimensions are needed to adequately represent human color vision . . . ." I take up these points in reverse order.

(a) The claim about additional dimensions beyond three fails to distinguish between the *intrinsic* dimensionality of the psychological space itself and any higher-dimensional space in which that psychological space may be embedded. If a subject can match any presented color of a specified type (whether color of surface or color of light) by adjusting just three parameters of a color-mixing device (and if, as in the typical human case, there are only three spectrally distinct classes of cones), then the intrinsic dimensionality of the color space must be three. But if that space has an intrinsic curvature, it may be usefully represented in a higher-dimensional (Euclidean) embedding space. Of course, it is an empirical question whether a given subject can match any color by adjusting just three parameters. It is likely that some animals (and, yes, possibly some human females) are genetically endowed with an additional spectral-sensitivity class (or classes) of cones. If such individuals are also endowed with the requisite additional neural circuitry, it is possible that they actually represent colors in a space of more than three intrinsic dimensions.

(b) The claim that MDS does not "tell you how to . . . measure distances" is, as it stands, simply wrong. The power of nonmetric MDS is that it is capable of converting a merely qualitative (for example, a rank order) scale of similarity into a quantitative (specifically, a ratio) scale of distances (see Shepard 1962a; 1962b; 1966; 1980; also Kruskal 1964a). (Perhaps the authors mean "measure distances on a physical scale" – but this would be, again, to conflate the physical and the psychological.)

(c) If "there is nothing [in the data] to tell you how to define the axes," then, obviously, such axes are irrelevant for the phenomena as captured by those data. In fact, colors are the canonical example of stimuli for which dimensions are very nearly perceptually "integral" and, to that extent, much less perceptually salient than such perceptually "separable" dimensions as size, orientation, and lightness (see, especially, Shepard 1991).

This is not to say that there are no perceptually discernable axes – including, perhaps, **Decock & van Brakel**'s "black/white, dark/light, and dull/bright" axes (in the color spaces appropriate for intrinsic surface colors, lighted appearances of surface, or lights themselves, as noted above). It is only to suggest that such psychological "axes" may have an importance secondary to that of the dimensionality of the space itself, which is definable without reference to any particular axes. On the other hand, if such axes are psycho-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

logically effective enough to affect the data, then MDS methods (such as those of three-way, individual differences scaling, or those that fit non-Euclidean metrics, such as Minkowski *r*-metrics) are in fact capable of recovering those axes (see, for example, Carroll & Chang 1970; Kruskal 1964a; 1964b; Shepard 1980; 1991).

## SR4. On generalization

### *SR4.1. Tenenbaum & Griffiths*

The article by Joshua **TENENBAUM** and Thomas **GRIFFITHS** presents an exciting and extremely promising extension of the theory of generalization presented in Shepard (1987b, as well as in my target article here). **TENENBAUM & GRIFFITHS**' article well makes its own case, (a) for the power of the general Bayesian approach to learning and generalization; (b) for the light that such an approach sheds on representations of similarity; and (c) for the importance of a "size principle" in Bayesian inference. I shall therefore confine myself to a brief statement of the essential ideas behind my original theory, and then, to some remarks about how I see TENENBAUM & GRIFFITHS' work as related to and as going well beyond my earlier work.

Although I did not emphasize this in the original presentation of my theory of generalization (Shepard 1987b), my formulation, like that of **TENENBAUM & GRIFFITHS**, is fundamentally Bayesian. The probability that a response will be made to any particular object is obtained by "hypothesis averaging," as TENENBAUM & GRIFFITHS put it. Each of the hypotheses over which the averaging is performed corresponds to a particular subset of the possible objects – namely, the subset that includes all and only the objects that are of a kind assumed (in that hypothesis) to harbor some significant consequence. The consequence might be a positive one in which each object in the subset might be found, for example, tasty and nourishing, or it might be a negative one, in which each object in the subset might be found, for example, capable of inflicting a painful bite or sting.

For objects that vary continuously in their perceptible qualities (such as size, shape, color, texture, manner of moving, and so on), each hypothesis can be regarded as specifying a corresponding, potentially *consequential* region in the continuous representational space. But, the "space" of the objects need not be continuous. Each hypothesis will still correspond to a subset of the possible objects – represented, for example, as the nodes of a discreet tree or graph-theoretic structure.

In order to derive a "generalization function" relating the probability that a response learned to one object will be made to another, however, the representational space or structure must provide a measure of distance between the two objects in every pair. A physical measure of distance cannot in general be expected to yield an invariant function. Invariance can only be achieved if the distance is the appropriate "psychological" distance. Perhaps surprisingly, the constraints inherent in distances (especially the *triangle inequality*) permit those distances to be uniquely and objectively determined – without circularity – from the generalization data themselves. In practice, this can be achieved by applying multidimensional scaling or related methods (of, for example, tree-fitting, graph-fitting, or "non-dimensional" scaling) to those data (Shepard 1980; 1987b). The exponential-decay form of the generalization function has been empirically revealed by plotting many sets of generalization data against the distances obtained in this way from those data.

Abstractly formulated in terms of the representational space, the desired probability of generalization was theoretically derived, in Bayesian fashion, as the ratio of two quantities. The numerator is the sum of the prior probabilities for all candidate regions in the representational space that contain both the point representing the object already found to be consequential, and the point representing the newly encountered object. The denominator is the sum of the prior probabilities for all candidate regions that contain the point representing the object already found to be consequential – *whether or not* that region also includes the point corresponding to the new object. This is the "hypothesis averaging" referred to by **TENENBAUM & GRIFFITHS**. The exponential manner in which the probability of generalization was computed to fall off with distance and, also, the computed *r*-metric form of the distances both turned out to be surprisingly insensitive to how the prior probabilities were assumed to depend on the sizes (and shapes) of the candidate regions (Shepard 1987b). (In agreement with empirical results for objects with "separable" or "integral" dimensions, however, the value of *r* that best characterized the metric of the representational space depended critically on whether the extensions of the candidate regions along the axes of the space were assumed to be correlated or uncorrelated – see Shepard 1987b; also, Myung & Shepard 1996.)

Of course, as an individual encounters additional objects and finds them to be consequential or not, the pristine exponential-decay form of the generalization function will be appropriately, and often markedly altered by Bayesian probability revision. (Such alteration was demonstrated, and shown to correspond to empirical findings in classification learning, in the early simulations reported by Shepard & Kannappan 1991, and by Shepard & Tenenbaum 1991.)

A very important way in which **TENENBAUM & GRIFFITHS** have gone beyond my original work on generalization is in considering the case in which an individual encounters objects belonging only to a particular consequential subset (their "strong sampling" case) and in introducing a corresponding "size principle," which weights hypotheses proportionally to the inverse of the size of the corresponding subset or region. This has enabled TENENBAUM & GRIFFITHS to derive a number of important results, including the empirically confirmed sharpening of "drop-off" of generalization around a consequential region within which more objects have been (randomly) encountered (see their Fig. 3) or on a dimension of the space along which the encountered objects fall within a more restricted range (see their Fig. 4).

An even more profound extension that Tenenbaum and his co-workers are exploring is the formation of representational structures (discussed in their commentary here) and, especially, approximations to continuous spaces and their geodesics (presented in Tenenbaum et al. 2000). The latter may help to show how learning, as well as evolution, shapes the kinds of spaces underlying the representation of transformations considered in section SR2 of my Response here. (It may also provide a way of constructing "neural spaces" such as **Edelman** discusses in his commentary, and may help to alleviate the concern that **Hoffman** expresses in his commentary – that, "Finding the ponderous multi-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

variate calculations of MDS internalized in actual brain tissue would be surprising indeed.")

In one respect, however, I believe that my original formulation of generalization was more general than **TENENBAUM & GRIFFITHS** suggest when they speak of stimuli, such as numbers, as not being "easily represented in strictly spatial terms." True, their Figure 5 shows an extremely nonmonotonic and spiky probability of "generalization" from the number 60, when plotted along the number line from 0 to 100. (This reflects the fact that generalization is, for example, much greater from 60 to other multiples of 10 than to closely neighboring numbers such as 59 and 61.) But this number line is not, of course, the representational space of numbers that would be obtained by multidimensional scaling of generalization data.

In the MDS solutions for just the concepts of the numbers 0 through 9, Shepard et al. (1975) obtained a representation in which, as **TENENBAUM & GRIFFITHS** note, the proximities of the points corresponding to these numbers consistently represented such common features as being even or odd, multiples of 3, and so on. So generalization might well decrease in an exponential way in the appropriate representational space. I return to this point toward the end of my following consideration of the commentary by **Chater, Vitanyi, and Stewart**.

### *SR4.2. Chater, Vitanyi, and Stewart*

Of two major issues raised in the commentary by Nick **Chater**, Paul **Vitanyi**, and Neil **Stewart**, the first concerns the mismatch they see between my theory of generalization and the experimental paradigms that yielded the data I have analyzed in support of the theory. Many of these paradigms might seem, as they evidently do to these commentators, to have more to do with *discriminatory confusion* than with *cognitive generalization,* as I define the latter.

I am in full agreement about the desirability of more data relevant to the simplest and purest case of generalization specified by the theory – that is, the case in which generalization is assessed by presenting a single new test stimulus following a single presentation of a novel training stimulus. Unfortunately, this is an extremely inefficient and impractical way to collect data. Moreover, in the artificial laboratory conditions under which human subjects are usually tested, this type of experiment, having little "ecological validity," is apt to leave subjects confused about what they are supposed to do. Such subjects may be inordinately affected either by their own interpretation or by variations in the experimenters' instructions. I wonder whether this might explain why **Chater et al.**'s own data led them to suggest that "generalization may be surprisingly variable, both between individuals and across trials, even with remarkably simple stimuli."

Until a better way of testing the theory is devised, however, I am not persuaded that the more readily available types of data that I have so far used are inappropriate for confirming the exponential-decay form of the generalization function. From my theoretical standpoint, there is this fundamental difference between generalization and confusion (that is, failure of discrimination). The form of the generalization function is determined by *cognitive uncertainty* about which – of all the variously sized, shaped, and located possible subsets of stimuli – is the consequential one. The form of the discrimination function, in contrast, is determined by *perceptual uncertainty* about the identities, or locations in the representational space, of the individual stimuli – independent of any hypotheses about the consequential subset.

The expected empirical consequence is that data that fall off exponentially with appropriate distances (for example, distances obtained by applying MDS to the data) should conform to the prediction of the (cognitive) generalization theory. Accordingly, such data do provide support for that theory, especially given that confusion (or failure of discrimination) – arising as it presumably does from random processes – is expected (except under special circumstances) to tend toward a function that has a Gaussian inflection (Nosofsky 1985; Shepard 1986; 1987b).

Admittedly, my caveat "except under special circumstances" is significant here. Such circumstances include those of continuing, repeated presentation with feedback – as in the "identification paradigm" described by **Chater et al.** – which have often been used in the psychological laboratory. I had long ago shown that under these conditions, the cumulative effects of spreading Gaussian memory traces of different ages also yield an exponential-decay of response probability with distance (Shepard 1958; 1986).

But it is by no means the case that all of the data yielding "generalization" functions approximating the exponential decay form have arisen from this paradigm. Examples include pigeons' rates of pecking a key illuminated by various spectral hues, without feedback, following intermittent reinforcement for pecking a key illuminated by a single training wavelength – as in the operant conditioning experiments by Guttman and Kalish (1956) and by Blough (1961). The results of my reanalyses of these data are displayed in Panels E and H, respectively, in Figure 1 of Shepard 1987b (also see Panel a of Fig. 10 of my target article). Moreover, other types of similarity data, which, I claim, (a) reflect the same internal representations as those that underlie generalization but (b) arise from stimuli that are not confused with each other, yield the same sort of exponential-decay function – as shown in Panels J and K in Figure 1 of Shepard 1987b (also see Panel b of Fig. 10 in my target article).

The second of the two major issues raised by **Chater et al.** concerns the limitation they claim to see on the range of types of stimulus objects or items to which my theory of generalization applies. I grant that the sets of data I used to provide evidence for the exponential-decay form of the generalization function were largely from experiments in which the stimuli were representable as points in some continuous psychological space with dimensions corresponding, for example, to visual size, shape, lightness, or hue; or to auditory pitch, duration, and so on. I also grant that the MDS methods I applied were mostly ones that correspondingly sought a representation in a continuous, usually Euclidean space. But the general theory (which could not be fully set forth within the length limitations of a *Science* article) is not, as the commentators seem to imply, restricted to "Euclidean distance in an internal multidimensional space."

Even in that original 1987b article, I explicitly deduced from the theory that for stimuli differing along perceptually "separable dimensions," the metric of the representational space is non-Euclidean – approximating one with a Minkowski $r$-metric with $r$ closer to the "city-block" value of 1 than to the Euclidean value of 2 (also see Myung & Shepard 1996). Subsequent analyses have provided additional



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

empirical confirmation for this deduction (Shepard 1991). Nor is there any requirement that the space be continuous. Stuart Russell (1988), for example, showed that the exponential generalization function is also deducible for the case in which the objects correspond to the corners of an $n$-dimensional Boolean cube (the case extensively investigated, for $n = 3$, by Shepard et al. 1961; also see Gluck 1991).

Indeed, I have proposed that the form of the generalization function is discoverable without assuming *any* particular form or metric for the psychological space. Given sufficient data, all that need be assumed is that the to-be-recovered psychological distances satisfy the metric axioms (most crucially, the triangle inequality), provided only that the variance of the distances (subject to this constraint) is maximized (Cunningham & Shepard 1974). (The maximization of variance, which is necessary for achieving a determinate solution, is analogous to minimizing the number of dimensions of a continuous space in the more standard methods of MDS.) Cunningham and I showed, for example, that this method can recover distances and, hence the generalization function, when those distances are path distances in an arbitrary tree structure. The generalization function shown in panel L of Figure 1 in Shepard (1987b) was in fact obtained by this method, without assuming any particular type of "space."

I remain to be convinced that the exponential generalization function cannot be demonstrated by applying such methods to objects or items of any kind – whether perceptual, conceptual, "part-whole structures," "scripts, sentences, or whatever" (to use **Chater et al.**'s terms). Although a somewhat different model is required to fully account for asymmetric generalization data, which are to be expected in some cases (see Tversky 1977), there is no reason why such data would not also be consistent with the exponential decay function. In short, I suggest that the law of generalization I have proposed may already be considerably more general than these commentators suppose.

At the same time, I hasten to add that I have great enthusiasm for the proposal of **Chater et al.** to relate generalization to a Kolmogorov-type measure of the complexity of the least complex process that will transform one such object or item into another. This could indeed represent a major step toward greater generality and understanding. Together with the possibilities being pursued by Tenenbaum et al. 2000, it might help to forge a deeper connection between inductive inference and the representation of spatial transformations (see sect. SR2 here; also Shepard 1997).

### *SR4.3. Dowe and Oppy*

I applaud the Komogorov-like Minimum Message Length (MML) approach mentioned by David **Dowe** and Graham **Oppy** for the same reasons that I do the related approach being taken by **Chater et al.** As I have just suggested in my response to the latter's commentary, however, I do not see these approaches as being inconsistent with my own Bayesian approach. I would nevertheless like to respond to some of the issues that **Dowe & Oppy** have specifically raised about my approach to generalization and about **TENENBAUM & GRIFFITHS**' extension of such an approach to more general problems of Bayesian inference.

My proposal that each of an individual's implicit "hypotheses" about what subset (or "basic kind") of things in the world might have a significant consequence corresponds, as **Dowe & Oppy** appropriately note, to "a connected local region in the space of possible objects." But I have some problems with their dismissal of such a correspondence as "not a 'fact about the world' at all" but "rather, an analytic or *a priori* truth which connects together the notions of 'basic kind' and 'connected local region . . . '," and with their implication that that they have somehow undermined my claim to having found a "universal law of generalization."

First, I take it as a plausible working hypothesis that any two objects of the same basic kind can be continuously transformed into each other without passing through an object that is not of that kind. But does the notion of basic kind exclude *a priori* the possibility that two objects (perhaps somewhat like the letters "d" and "D") could be considered to be of the same basic kind even though every continuous deformation of one into the other must pass through something that is not recognizable as being of that kind?

Second, even if the correspondence between basic kinds and connectedness in the representational space were true *a priori*, that in itself would not seem to exclude the possibility of worlds in which the concept of basic kind simply has no application. Consider, for example, a very different world from ours in which all "things" – including any "consequences" – grade continuously into everything else with no distinguishable boundaries or gaps. Moreover, as I have indicated earlier in responding to the commentators, I interpret the phrase "facts about the world" very broadly, to include facts of mathematics and logic, including geometry, set theory, and probability.

Finally, it turns out that "connectedness," as such, is not critical for deriving the exponential-decay generalization function. Derivation of the law of generalization was facilitated by making some specification of the possible shapes of "consequential regions" – such as specifying they are connected and/or convex. But numerical explorations indicate that the theoretically required Bayesian integration over all candidate regions of whatever form is specified is remarkably insensitive to such specifications, uniformly yielding the exponential-decay law for any reasonable specification (and approximating the same class of Minkowski $r$-metrics). The exponential function and the metric appear to be essentially "invariant" not only when there are changes in assumptions about the shapes of consequential regions but also when there are minor disconnections of those regions – at least if the disconnected islands are still reasonably close to a localized center.

With regard to the fundamentally Bayesian basis of my theory of generalization and of the important extensions proposed by **TENENBAUM & GRIFFITHS**, **Dowe & Oppy** object that "we know from countless experiments on people that we are very far from being perfect Bayesian reasoners . . . " The problem here is that Dowe & Oppy neglect to distinguish between (a) peoples' conscious reasoning about statistical matters as reflected, for example, by their performance on paper-and-pencil tests, and (b) the implicit processes of probability revision that goes on at deep and consciously inaccessible levels of peoples' perceptual/representational systems. Performances on tests of the former type, which have challenged humankind only very recently on the evolutionary time line, are indeed "very far from being perfect." In contrast, already extensive and increasing evidence indicates that our unconscious processes of per-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ceptual inference, which have been shaped over a vastly longer evolutionary history, may closely approximate Bayesian norms.

### SR4.4. Lomas

As should be clear from the remarks I have just made in response to **Dowe & Oppy** concerning the degree to which our perceptual systems approximate optimum, Bayesian principles, I cannot endorse Dennis **Lomas**'s statement that "perceptual mechanisms, which are more likely [than the more rapidly changing 'belief systems'] to involve evolutionary internalization of universal regularities, are *unreliable*" (emphasis mine).

The principal difficulty I have with the commentary by **Lomas**, however, is that he evidently takes me to be basing my theory of generalization on the assumption that the specific "basic kinds" that we living humans find around us are universals. I make no such assumption, nor does my derivation of a proposed universal law of generalization require any such assumption. What I *do* assume and use for the derivation is only that biologically significant objects universally do belong to basic kinds – without assuming anything further about what specific kinds may exist at any given time. Hence, his observation that: "Animal species arise, decline, and disappear, others arise, and so on," is a truism that is irrelevant to the question of universality.

Incidentally, I don't agree with **Lomas** that categorization of something as a dog, as opposed to a statue of a dog, "is beyond the reach of perceptual capacities." If it were, we would not be able to discriminate dogs (which belong to a natural kind with its own distinct consequences or affordances) from statues of dogs. A real dog is perceptually discriminable from even the visually most life-like three-dimensional statue from the way it jumps about, barks, wags its tail, sniffs, licks, slobbers, and smells. As Lomas surmises, I do hold, however, "that the cognitive resources [for categorization] are not restricted to perceptual resources." Certainly we recognize that something is a tool, even though there is no single perceptual form or feature that distinguishes all tools from all other objects. Objects thus abstractly categorized by function may nevertheless be represented as potentially "consequential sets" and, thereby, contribute to the mediation of generalization. Representations of such objects, even if not members of a connected region in a continuous space, may be connected by links in a graph-theoretic structure (see Pruzansky et al.1982).

### SR4.5. Movellan and Nelson

I second the recommendation of Javier **Movellan** and Jonathan **Nelson** to put aside endless debates about "undecidable structural issues . . . in favor of a rigorous understanding of the problems solved by organisms in their natural environments." I have, however, this reservation about their suggestion of the term "probabilistic functionalism" as "a unifying paradigm for the cognitive sciences": That term does not adequately describe my own approach to understanding the representation of objects, their orientations, spatial transformations, and colors, for, (as I argued in the preceding sects. SR1–3) my approach is not fundamentally probabilistic. (It is, as I noted in my response to **BARLOW**, more Gibsonian than Brunswikian.)

**Movellan & Nelson** question my representation of hypotheses as sharply bounded subsets to which any given object must categorically either belong or not. As they note, Rosch et al. (1976) and others have amply demonstrated that members of a category differ in how psychologically representative they are of that category (thus, for people, a robin is more representative than a penguin of the category "bird"). Certainly, this factor of representativeness or typicality is psychologically significant. It manifests itself in many ways, including the time required to verify whether a presented item is a member of the category, the order in which items are freely recalled from the category, and so on. But this should not obscure the fundamental fact that except when forced to respond with extreme rapidity, people are virtually 100% accurate in classifying a penguin as a bird. Again invoking Chomsky's (1965) distinction, peoples' classificatory *competence* may not be fully revealed by their *performance* under time pressure.

It is completely compatible with my general Bayesian approach to admit hypotheses that are not sharply bounded. I (like Anderson 1990) have in fact considered hypotheses corresponding to functions with inflected, Gaussian-like boundaries, rather than to purely "binary membership functions" in the representational space. Indeed, numerical explorations indicated that (in the absence of differential reinforcement) integration over Gaussian functions with a suitable distribution of variances yields a concave-upward function resembling, again, the exponential-decay generalization function. Certainly, this is more like the function shown in the upper panel of **Movellan & Nelson**'s Figure 1, than like the lower one. (A strict exponential decay function was derived by integration over Gaussian distributions under somewhat different conditions in Shepard 1958.)

My own principal motivation for considering such graded consequential distributions arose from the consideration that although all objects of a given basic kind may have the potential for a certain significant consequence, the probability of manifesting that consequence may be less for marginal exemplars. (A certain kind of fruit may be edible, but less tasty or nourishing if it is too small and green or, at the other extreme, too dark and withered.) Nevertheless, I chose to begin by exploring how much could be explained by hypotheses of the latter, sharply bounded type for the following two strategic reasons.

First, an individual whose hypotheses all correspond to sharply bounded regions would, under Bayesian probability revision, quickly converge on the correct hypothesis if that hypothesis corresponds to a consequential region that does have sharp boundaries. The resulting minimization of erroneous responses would confer a significant advantage for that individual over individuals whose hypotheses correspond to fixed Gaussian consequential distributions like those that **Movellan & Nelson** assumed in deducing the function exhibited in the lower panel of their Figure 1. The resulting, irreducibly inflected and tapered, generalization function would entail that the individual continues forever to make erroneous decisions about objects in the vicinity of a consequential boundary that is, in fact, sharp.

Second, even if the correct hypothesis is not itself sharply bounded but has some other, perhaps Gaussian-like form (resembling, perhaps the one shown at the bottom of **Movellan & Nelson**'s Fig. 1), a properly weighted combination of overlapping and sharply-bounded candidate regions could approximate the correct Gaussian-like distribution.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

(Though, admittedly, what learning process will achieve such a convergence most effectively remains to be specified.) Conversely, of course, a sharply bounded region can also be approximated by integration over weighted Gaussian functions – but only if the variances of those functions are suitably distributed, rather than having a fixed value as in **Movellan & Nelson**'s example. But when the consequential region is in fact sharply bounded, convergence to that sharply-bounded region will almost certainly be slower when the hypotheses are all Gaussian than when the hypotheses are all sharply-bounded. It may even be, as I suspect, that convergence of Gaussian hypotheses to a sharply bounded region will be slower than convergence, inversely, of sharply-bounded hypotheses to a Gaussian distribution.

#### SR4.6. Cheng

Ken **Cheng** has nicely highlighted what I regard as a central aspect of my whole approach. This is my attempt to discover the *functional grounds* of general psychological principles, rather than being satisfied with merely describing empirical regularities in behavior, or even with identifying neural mechanisms that lead to those behaviors. This is indeed, as I stated in my introduction, to pursue psychology's fundamental "Why" questions.

In addition, I am indebted to **Cheng** for providing the first evidence for a concave-upward generalization function consistent with my proposed universal exponential-decay function in an invertebrate species – namely, the honeybee (Cheng 2000). As **Cheng** implies, evidence that this law may govern the behavior of such a distant and ancient species raises the question of how these animals "come up with" a principle such as I derived from a "cognitive" theory based on "hypotheses" corresponding to candidate consequential regions in representational space. It may be, as Cheng suggests, that these animals have somehow internalized an approximation to the theoretically optimal exponential function without generating this function by anything corresponding to a process of integration over hypotheses. If, however, further studies show that the form of the function is modified with differential reinforcement, as demonstrated in the connectionist simulations of Shepard and Kannappan (1991), or with presentation of multiple positive instances, in accordance with the extended Bayesian theory described here by **TENENBAUM & GRIFFITHS**, we may have to raise our estimates of the cognitive capabilities of the lowly bee. (As **Lee** acknowledges in his commentary, we do not suppose that even humans carry out anything resembling a symbolic integration over all hypotheses. Rather, we imagine generalization to arise from some more automatic and unconscious analog process, perhaps resembling that embodied in the simple "neural-net" model explored by Shepard & Kannappan 1991.)

#### SR4.7. Lee and Pothos *(two separate commentators on* TENENBAUM & GRIFFITHS*)*

The points concerning the need for "complexity measures" and the importance of "context effects" raised in the separate commentaries by Michael **Lee** and by Emmanuel **Pothos**, respectively, are worthy of discussion. I am hopeful that these points can be successfully addressed within the general Bayesian framework adopted by me and by **TENENBAUM & GRIFFITHS**. But (apart from my brief preceding reference to **Lee** in my response to **Cheng**), I shall leave it to Tenenbaum & Griffiths, to whom these two commentaries are specifically addressed, to provide responses.

### SR5. On my general approach to psychological science

#### SR5.1. Edelman

So far, my approach to finding "an invariant law" has been "predicated," as Shimon **Edelman** well puts it, "on the possibility of finding an 'abstract space' appropriate for its formulation." Beginning with my response to **BARLOW**'s article, I have tried to explain why I have so far given this endeavor priority over an attempt to discover how such an abstract space may be concretely represented in the brain. Yet, as **Edelman** implies, it must have some sort of isomorphic instantiation there – though one closer to my "second-order" isomorphism (Shepard & Chipman 1970; Shepard et al. 1975) than to a direct "first-order" isomorphism. In any case, the very impressive work that Edelman has already accomplished on representation (see Edelman 1999, and his suggestive commentary here) gives me hope that significant progress may now be possible toward plausible and useful formulations of "neural spaces." (Also relevant may be the approach to constructing representational spaces being explored by Tenenbaum et a1. 2000, to which I referred in my reply to **TENENBAUM & GRIFFITHS** in sect. SR4.)

#### SR5.2. Pribram

My former Stanford colleague Karl **Pribram**, as a neurophysiologist, understandably feels even more acutely my neglect of the concrete brain mechanisms that must underlie the representations and principles I have been formulating at an abstract – one might say "disembodied" – level. I was interested to see that he begins by specifically noting, quite independently of what I have here written in my introduction, that he credits me with providing some answers to questions about "what" the perceptual-cognitive process "is about" but with neglecting questions about "how" the process is physically "implemented." Thus, I have been dealing with only "half" of the problem. I, of course, claim that there are really three types of questions and, as I remarked in my response to **BARLOW**, that I have been most basically preoccupied with the third – the "Why" questions. It is primarily through answering the "Why" questions, I suggest, that we can gain the most useful guidance for addressing the "How" questions.

I was intrigued by the evidence, cited by **Pribram**, that Grossman and Martha Wilson have reported for neuroanatomically distinct locations for representing "image and object-form driven" categories such as "hue and shape" and the "comprehension-driven" categories such as "fruit and vegetables." These two types of representation have been found to be best fit by continuous spatial representations and by discrete tree-structures, respectively (see Pruzansky et al. 1982).

#### SR5.3. Dresp

It should by now be clear that in searching for invariant psychological laws, I do not accept Birgitta **Dresp**'s implica-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

tion that everything in the environment is "steadily changing." The environmental features on which I have principally focused in my target article – namely, the 24-hour period of the circadian cycle, the *principles* of conservation of angular momentum and of mass, and, especially, the three-dimensional, Euclidean character of physical space, as well as the *fact* that significant objects belong to distinct kinds – are not changing at all. Of course, other things do change, and for those we need "learning" and "probabilistic processing" (which figure prominently in the topic of generalization discussed in my preceding sect. SR4).

But do the principles of learning and of probabilistic processing that are most adaptive in the world themselves change? Without at all wishing to deprecate Grossberg's Adaptive Resonance Theory of the brain (ART), which **Dresp** forwards as "a more powerful alternative" that offers "mechanisms" for the explanation of "representational activities," I would ask Dresp to consider the following: If ART or any other such theory is effective in this way, do not the principles that that theory embodies reflect some unchanging facts about the otherwise "steadily changing" world? And, to the extent that these principles do reflect such unchanging facts, might not the principles be regarded as internally representing those facts?

In any case, I do not hold that the "concept of internalization implies that the brain stores multiple copies of objects and events and all their possible relations in space and in time," as **Dresp** puts it. It is in fact one of the major advantages of an internalized capacity for representing (geodesic) transformations of objects in space that the internal representation of any particular position of an object may be realized by iteration of a very small transformation on a representation of the object in some canonical position, thus wholly eliminating the need for storing all possible positions of all possible objects.

### SR5.4. Gerbino

We move, now, from contemporary neurophysiologists back to the Gestalt psychologists, who had a particular view of the "dynamics" of the brain. As **Gerbino** notes, I have shared with the founders of the Gestalt school, Wertheimer, Koffka, and Köhler, an interest in the role of minimizing principles in perceptual organization and in what the experience of apparent motion, in the absence of any physically presented motion, can tell us about the organizing principles governing perception and representation (which, certainly, are implemented in the brain). Yet, the view to which I have come is, it seems to me, fundamentally different from that put forward by these Gestalt psychologists.

For the Gestaltists, it seems, the brain manifests inherent minimizing principles simply by virtue of being a physical system – much as a soap bubble assumes (in Köhler's familiar example) the "good form" of a sphere by minimizing the surface area enclosing a conserved volume of air. Gerbino quotes Koffka (1935) as stating that "A process must find its explanation in the dynamics of the system within which it occurs; the concept of biological advantage . . . does not belong to dynamics at all." In contrast, my more evolutionary view is that the "grey matter" of the brain, unlike matter in general, has been specifically shaped to provide a veridical representation of what is going on in the external world. For me, symmetry and minimization principles are of fundamental importance. But they come into play not as concrete manifestations of the necessarily material constitution of brain stuff but, rather, as reflections of very abstract mathematical (and especially geometrical) features of the external world that have been specifically "internalized." Granted this internalization must have some physical embodiment in the neuronal machinery of the brain, but it is an embodiment shaped by natural selection and subsequent fine tuning through individual learning.

### SR5.5. Whitmyer

In comparing my views to those of the Gestalt psychologists, Virgil **Whitmyer** (like Karl **Pribram**) wants to shift what he takes to be my preoccupation with the question of "*what*" is represented to the question of "*how*" it is represented "in neural stuff." But, once again, my primary concern throughout has been with what I regard as the deeper question of "Why" psychological laws have the forms that they do.

In addition to his focus on the brain, **Whitmyer** also brings up something else that concerned the Gestalt psychologists – namely, the "phenomenal states" presumed to accompany or to arise from the processes in the brain. Whitmyer expresses uncertainty about whether, in talking about internalization, I am concerned with the "mapping" from "neural elements" to such phenomenal states, or to what is represented (largely in the external world). Actually, since I am not dealing with neural elements at all, I am not concerned (here, anyway) with *either* of these two mappings.

With respect to the issues that **Whitmyer** raises, I am concerned, rather, with two different things: The first is the characterization of the problem that a perceptual, cognitive, and mobile individual faces in the world (including identifying the accessible invariants – as in Gibson's "ambient optic array"). The second arises from the experimental work that my students and I have reported on apparent motion, imagery (including mental rotation), and generalization. It is the (quite non-Gibsonian) demonstration that principles relevant to solving the problems faced in the world are indeed internally represented, even when the relevant sensory information is absent. Our work demonstrates, I believe, that general laws can be discovered in this way without knowing anything whatever about their physical implementation in the brain or, indeed, without any direct access to the phenomenal states of our experimental subjects.

This is not to say that I am not interested in the deep and puzzling issues about consciousness – mind versus matter, the mental versus the physical, the relation of *qualia* to brain states, and the problem of free will. I very much am (see, e.g., Shepard 1993). The *scientific* matters with which I have been dealing here do not, however, depend in any way on the resolution of these difficult and contentious *philosophical* issues. A person who is looking at a "circular structure" in the world will report the phenomenal experience of a circle but, of course, as **Whitmyer** notes, "the neural version won't look circular to an observer gazing at the cortex [of that person]." This can be dealt with scientifically, by invoking what I have referred to as "second-order" isomorphism (Shepard & Chipman 1970; Shepard et al. 1975). For this, I need to assume only two things: (a) There is some unique pattern of activity that takes place in the



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

brain when and only when the person perceives or (as I also allow) dreams of, hallucinates, or even just imagines a circle. (b) The pattern has some overlap (or elements in common) with a pattern that arises when and only when the person experiences a similar object (such as a many-sided regular polygon, an ellipse, or a sphere) – with the extent of overlap corresponding to the degree of experienced similarity. I do not need to assume or to know anything else about that pattern of activity or its neuronal character – and, certainly, not that it achieve the "first-order isomorphism" of being itself literally circular.

### SR5.6. Zimmer

Concerning perception as phenomenal experience, Alf **Zimmer** raises the question of how it can be "veridical" if the external reality (Kant's numinal realm of *Ding an sich*), about which we would like to think our perceptual experience can inform us, is categorically different and irreducibly "transphenomenal." This is again a deep and, to me, fascinating issue relevant for the philosophically inclined physicist or psychologist. But it is one whose resolution is not, I believe, essential for the program of scientific research I have been discussing here.

For the purposes of this program, I am content to hold to the answer I have proposed to Koffka's question, "Why does the world appear the way it does – is it because the world is the way it is or is it because we are the way we are?" (Koffka 1935). My answer (which I am grateful to Zimmer for quoting, because it so succinctly expresses my view on perception) is this: "The world appears the way it does because we are the way we are; *and* we are the way we are because we have evolved in a world the way it is." (I had not, however, noticed that the date of publication of my proposed answer followed the publication of Kant's *Critique of Pure Reason* [1781/1968] by exactly 200 years. *Vielen Dank!* Alf.)

### SR5.7. Mausfeld

My view of perception, as represented in my just-quoted answer to Koffka, might indeed justify Rainer **Mausfeld**'s classification of me as being (along with Herbert Spencer) a "nativistic empiricist." I am, however, uneasy about some other ways in which **Mausfeld** describes my position. For example, his characterization of my "claim that there is an 'evolutionary trend toward increasing internalization'" as "non-Darwinian" is, I suspect, based on an understandable misinterpretation of what I intend by such a claim. I did not mean to imply, by the word "trend," that (in addition to random variation and natural selection) there is some occult force favoring internalization. I merely meant that such a process of random variation with selection – beginning, as it must, with the simplest possible living forms – will inevitably lead to some forms that achieve adaptations to more and more complex features of the environment. As I pointed out (in the article that Mausfeld cites in this connection – Shepard 1987a), this is directly analogous to the fact that among the particles embarked on purely random walks from a starting location next to a barrier, some will inevitably be found farther and farther from that barrier.

I am also uncomfortable with **Mausfeld**'s characterization of what he finds even "more problematic" – namely, what he refers to as "Shepard's extreme physicalistic stance." If by a "physicalistic stance" he means one that denies the reality of phenomenal experience or qualia (which, I admit I have not explicitly dealt with in the present *scientific* context), he will find that in other, more *philosophical* contexts, I have made clear that I am by no means a physicalist (see, for example, Shepard 1993; 1995b; Shepard & Hut 1998 – see, also, my final sect. SR6.3, here). Additionally, as I have emphasized, the "world" from which I claim that facts may become "internalized" includes abstract mathematical as well as concrete physical facts. This may also help to counteract **Mausfeld**'s imputation that I assume "that the rich structure is imprinted on the mind of the perceiver almost entirely from without." Anyway, if he has some *non-arbitrary* source for mental *structure* – beyond physical reality and mathematical (including, for example, game-theoretic) facts – I should very much welcome fully detailed instruction about the nature of that source.

### SR5.8. Kurthen

It is not, I think, an inordinate charitableness and humility that has prevented philosopher Martin **Kurthen** from making plain what he means by a feature of the world "being *literally* represented in the organism" (emphasis mine), and exactly how he regards this as differing from the organism behaving "*as if it had internalized*" that feature (emphasis his). The horrendous implication of such a distinction (already stressed by **KUBOVY & EPSTEIN**) continues to elude me. Nevertheless, I can sense the somewhat Gibsonian appeal of **Kurthen**'s Heideggerian "radical ecological embodiment." (I have the image, here, of an organism and the ecological niche within which it resides, as a system that resonates as a whole.) In any case, I like to think that those of us endeavoring to build a psychological science will be open to any *scientific* contribution that a radical embodiment approach may be shown to offer.

### SR5.9. Pickering

We come now to commentators who raise questions about the ways in which implicit knowledge about the world may become internalized. I accept John **Pickering**'s point that "the internal basis of cognitive-perceptual skills is likely to blend ontogenetic and phylogenetic learning" occurring "over different time scales," and is likely to do so, particularly, as it interacts with the "cultural scaffolding that surrounds human development." I also agree that "learning influences both the speed and the direction of evolutionary change" (as in "the Baldwin Effect"), and that our own cultural and material products change the environment in ways that must increasingly exert some evolutionary selection on human (and other animal) lines. I am inclined, however, to think that **Pickering** may overestimate the extent and speed of resulting evolutionary internalization. The rate of global technological and cultural change seems, rather, to be outstripping the ability of "natural" selection to keep up. There are many widely noted indications that we may actually be less adapted to (and less "at home" with) the rapidly shifting challenges of the modern world we are in the process of creating than our pre-technological ancestors were to the more stable demands of their Pleistocene world.

At the same time, however, the *explicit* knowledge about the world that some humans have been gaining through sci-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ence just within the last few hundred years goes far beyond any increment in our biologically internalized *implicit* knowledge. From now on, the further knowledge that humans are gaining about the world, evidently at an ever-increasing rate, may be represented little, if at all, in our genes. Rather, it may be almost entirely represented in what (following Dawkins 1999) might be termed our "memes." This development, apparently unprecedented among terrestrial species, has enormous implications – to which I shall return in my final section SR6.

### SR5.10. Heschl

Adolf **Heschl** is of course correct in pointing out that in addition to any approximations to perceptual-cognitive universals, evolution has given rise to the "fascinating . . . diversity" of the life we find on Earth. But he goes on to say that my "perfectly reductionistic and behaviorist" thesis, as he curiously describes it, "is clearly contradicted by anything which has been brought to light by evolutionary biology since its very beginning with Charles Darwin." In the first section of my introduction, I tried to explain how Darwinian evolution could in fact produce a converging "mesh" of the mind to the world, despite the marked anatomical diversity such evolution has produced. I argued that while diverse bodily forms are capable of functioning in the world, there is a selective pressure to represent the physical and mathematical facts of the external world as they exist independently of our own bodies, thus tending to close a "psychophysical circle."

### SR5.11. Raffone, Belardinelli, and van Leeuwen

What I have just said in my response to **Heschl** applies equally to the commentary of Antonino **Raffone**, Marta **Belardinelli**, and Cees **van Leeuwen** – particularly to their assertion that "the structural and dynamic properties of the sensory and effector systems . . . may significantly differ between organisms in the universe, due to varying ecological constraints" and, hence, that "Shepard's notion of evolutionary internalization may not be a plausible general bio-cognitive principle." I should perhaps add, however, that I do not claim that there is an "adaptive value of *contemplating* the universality, invariance and elegance of the principles which govern the external world" (emphasis theirs as well as mine!). Nor do I assume "that the mind *performs mathematical calculations* in terms of spatial coordinates or static representations" (emphasis mine). As I have remarked before, I hold that the implicit mental processes that guide our daily lives (as opposed to the explicit mathematical calculations of the scientist) are of an analog nature. (I further discuss these two kinds of mental processes in the ensuing sect. SR6.)

### SR5.12. D. Schwartz

I wholeheartedly embrace the possibility proposed by David **Schwartz** that "nonarbitrary principles may fruitfully be sought not only in the laws of physics and mathematics, but also in the logical entailments of different categories of representation." Without having explicitly dealt with indexicality, I had hoped that such things as "logical entailments" might be provided for under the general rubric of mathematics and logic. I very much look forward to seeing what Schwartz and others are able to develop in the way of a detailed theory of indexical representation.

### SR5.13. Massaro

I thank Dominic **Massaro** for his very generous and supportive remarks. Clearly I need all the help I can get – and his is particularly welcome.

The fundamental question that **Massaro**'s work has addressed arises because information about the world comes in through very different, essentially incommensurate sensory modalities. Despite the diversity of its visual, auditory, or tactile forms, however, all this incoming information pertains to the same external world. Although I have not specifically discussed the question of the optimum way of combining information of such diverse forms, I would hope that the probabilistic part of the answer would be compatible with two general types of principles that I have discussed. The first of these is the Helmholtzian principle that perception represents that external situation, object, or event that is both (a) consistent with the sensory information and (b) most probable in the world (as I discussed in sect. SR2). The second is the principles of Bayesian inference (discussed in sect. SR4).

Oden and Massaro's "Fuzzy Logic Model of Perception" (FLMP) appears to be very much in the spirit of the general Bayesian approach advocated here. It evidently makes a simplifying independence assumption, yet it has proved to offer a remarkably effective approach to this problem of combining information from different sensory channels. Particularly influential has been the impressive demonstration by Massaro and his coworkers of the power of this approach in elucidating how people combine visual input from the motions of a speaker's face with auditory inputs from the speaker's voice in the understanding and localization of speech (see Oden & Massaro 1978).

Incidentally, as in the case of other perceptual-cognitive principles, the interesting question arises here whether principles of cross-modal integration may be partially innate (rather than entirely learned through experiential association, as Helmholtz's empiricism would seem to suppose). There are some striking indications that information obtained through one modality may immediately generalize to an entirely different modality, in the absence of any relevant prior experience.

One example is the ability, reported by Meltzoff and Moore (1977; 1999), of neonates to imitate facial expressions such as the opening of the mouth or the protruding of the tongue. Although the conclusive demonstration of this ability in the newborn infant is beset by methodological difficulties, the suggestion of such an ability raises an intriguing question: How might newborns already have an association between the *visual pattern* of another's facial expression and *motor act* of forming the same facial expression – given that their own face is not visible to them? Two other, less controversial examples are the ability of blindfolded human adults to recognize the identity of a letter or number (a) that is traced out in space by a spatial pattern of beeps sequentially emitted by individual small loudspeakers in a spatial array facing the subject (Lakatos 1993), or (b) that is traced on a patch of skin never before used for alphanumeric recognition (as in Joseph & Kubovy's 1994 experiment that I discussed in my commentary on **KUBOVY & EPSTEIN** in sect. SR2).



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

### *SR5.15. Bedford*

My discussion (in sect. SR2) of the role of apparent motion and mental rotation in establishing the identity of shapes presented in two different positions in space was confined almost exclusively to positions experienced within one sensory modality and, in the case of apparent motion, experienced sequentially. In most of the studies considered, the presentation was exclusively visual (and through both eyes), but some studies (mentioned in my reply to **Foster**) compared apparent motion in the different visual, auditory, and tactile modalities (Lakatos & Shepard 1997a; 1997b). As Felice **Bedford** notes, however, the "correspondence" problem of establishing object identity is much more general than this. It includes cases in which the information arising from the different positions comes in not only from "different times" and "different spatial locations," but also via "different modalities, . . . and even different eyes." Moreover, the object may be nonrigid, in which case the identification of two views as being of the same object cannot be established by finding a rigid transformation to full congruence.

For such cases, **Bedford** invokes the "whole set of nested geometries" of Felix Klein's famous *Erlanger program* in which "the familiar Euclidean geometry is only the beginning" of a series that goes on to include the successively more abstract geometries that are invariant under the correspondingly more general similarity, affine, projective, and topological transformations (see Klein 1893/1957). I am in complete sympathy with **Bedford**'s proposal, which appears to be in the spirit of my own interpretation (discussed in my reply to **BARLOW** in sect. SR2) of the emergence of nonrigid apparent motion as a reflection of a "hierachy of criteria of object identity" (see also **Foster**'s commentary and Leyton 1992).

Granted, I insist that the space in which we have evolved is three-dimensional and Euclidean – that is, a space that affords, for rigid objects, the translations, rotations, and screw-displacements that I have described for the Euclidean group. But this same Euclidean space also admits nonrigid transformations. These include: (a) the similarity transformation of a balloon inflating or the shadow of a planar object that is approaching or receding from its nearby source of illumination (also the retinal projection of such an object that is approaching or receding from the eye); (b) the affine shearing of a deck of cards (or of the projection of any distant tilting planar object, such as a circle into an ellipse, or a square into a rectangle or a parallelogram); (c) the projective transformation of the shadow of such a planar object induced by its rotation when it is close to the point-source of its illumination (or of the retinal image of the object itself when close to the eye); and, finally, (d) the topological deformation of anything from a lump of wet clay to an animal's body. So, yes, **Bedford** is correct in surmising that I "would agree" to "removing the restriction of Euclidean geometry." Indeed, I never intended such a "restriction" to apply to the representation of nonrigid transformations.

### *SR5.16. Todd & Gigerenzer*

As already indicated in my response to **Jacobs et al.** (sect. SR2), I have great respect and admiration for the demonstrations that Gerd Gigerenzer and Peter Todd (1999) have provided for the "satisficing" effectiveness of "fast and frugal heuristics." Indeed, I myself was an early advocate of the heuristic of making decisions based on one factor (or a simple, linear combination of just a few) rather than mentally laboring to achieve the optimum, generally nonlinear combination of all relevant factors (Shepard 1964c; compare Martignon & Hoffrage, pp. 119–40 in Gigerenzer & Todd 1999b). Nevertheless, my present search for universal perceptual-cognitive principles might seem antithetical to the approach represented here by **Todd & Gigerenzer**, with its "adaptive toolbox" reminiscent of Ramachandran's (1990a) "utilitarian," "bag-of-tricks" theory of perception. I therefore hasten to explain how I would reconcile my approach with that expressed by **Todd & Gigerenzer** – as both have in part grown out of an evolutionarily informed starting point.

First of all, I never claimed that the "universal laws" I have been proposing, are perfectly instantiated or "internalized" in any human being, let alone in all living organisms. What I have had in mind, rather, is that with the evolution of any beings exploiting the general "cognitive niche," natural selection tends to favor an increasing match of their internal representations to the general features of the world represented. At any given stage of evolution, development, and learning, the approximation may be implemented as a collection of heuristics – a kind of "bag of tricks" – that to an adequate degree approximates the optimum functionality. But, some continuing selection for still better approximations might reasonably be expected in cognitively competitive evolutionary lines.

Moreover, regardless of how far we humans have so far advanced along our own evolutionary path, I believe it useful to provide as much as we can in the way of an analysis of the problems posed by the world and a characterization of optimal principles for their solutions. Only in this way will we have an objective, ideal benchmark against which to evaluate either what may be theoretically proposed as candidate heuristics or what may be empirically found to be operating in any existing organisms, including ourselves.

In "Extending an ecological perspective to higher-order cognition," **Todd & Gigerenzer** argue that the satisficing "scissors" of Herbert Simon (which cut off from consideration aspects of the world that are not fully analyzed or wholly necessary for many situations; Simon 1990) "may be a better model . . . than Shepard's mirror" (presumed to provide veridical internal representations of all relevant aspects). Todd & Gigerenzer recommend this when one is "studying a range of mental mechanisms, *particularly higher-level ones*" (emphasis mine). In responding, I find it useful to distinguish not just lower-level and higher-level cognition, but *three* quite different levels of cognition. I suggest that it is the intermediate level (more than the highest level) that depends most extensively on fast and frugal heuristics. As I shall now argue (in the concluding sect. SR6), it is at the lowest and the highest levels of cognition that universal features of the world may be most faithfully (if always incompletely) reflected. The way in which knowledge about the world is internally represented at these two extreme levels is, however, totally different.

## SR6. New directions: Toward cognitive grounds for science and of ethics

Some seven years have elapsed since my target article was originally published (in Shepard 1994), and still more years since I advanced many of these ideas in earlier versions (as


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

in Shepard 1981b; 1984; 1987a; 1987b; and Shepard & Cooper 1982). Although the intervening years have given me no reason to retract my endorsement of the basic ideas, feedback has helped me to clarify and to sharpen their expression. I am particularly grateful to the many thoughtful researchers who have here honored me with probing critiques. I hope that my response (in the preceding sects. SR1–5) will have been found to go at least some way toward meeting some of the objections raised.

Since the first publication of my target article, I have also embarked on explorations of the implications of my approach for more far-reaching issues concerning humankind, rationality, and the cognitive grounds of science and of ethics. I deem it appropriate, in closing, to give some indication of the direction of these more recent explorations. I have two reasons for doing so: (a) These explorations have grown directly out of the ideas discussed here. (b) Yet, the book (*World and Mind*) in which I have been endeavoring to set forth these new directions is still in preparation. (In the meantime, parts of this still evolving material have, however, been presented in a number of lectures – including the Paul M. Fitts Memorial Lectures at the University of Michigan in 1993, the William James Lectures at Harvard in 1994, the Charles M. and Martha Hitchcock Lectures at the Berkeley and Santa Barbara campuses of the University of California in 1999 and, most recently (and succinctly), the Carl I. Hovland Lecture at Yale in 2000.)

### *SR6.1. The step to rationality*

I ended my responses to the individual commentators by remarking, in reply to **Todd & Gigerenzer**, that instead of just a lower and a higher level of cognition, I distinguish three fundamentally different levels. So far, I have almost exclusively referred to the lower two – particularly the lowest, which is the level of the most deeply internalized, implicit, automatic, and evolutionarily perfected representations of biologically relevant facts and principles of the world.

I believe my intermediate level may correspond to **Todd & Gigerenzer**'s "higher-level" cognition. It is the level wherein people consciously struggle with challenges that have confronted humankind for too short a time to have evolutionarily shaped mechanisms of the effectiveness, speed, and accuracy of those long operating at the lowest level. These recent challenges (mostly consequences – whether intended or unintended – of humankind's own interventions) include the paper-and-pencil tests that have been taken to indicate how far we fall short in our understanding and performance of logical deduction (e.g., Wason & Johnson-Laird 1972; Woodworth & Sells 1935), of probabilistic inference (e.g., Tversky & Kahneman 1974; 1981; 1983), or of reasoning about physical processes (e.g., McCloskey 1983; McCloskey et al. 1980; Proffitt & Gilden 1989; Proffitt et al. 1990). For example, as I noted in responding to **Dowe & Oppy** (in sect. SR4), their statement that people "are very far from being perfect Bayesian reasoners," was based on the conscious reasoning tapped by such "paper-and-pencil" tests rather than on evidence concerning the more evolutionarily perfected automatic processes of perception and cognition. Possibly **Todd & Gigerenzer** would agree with my claim that tests designed to engage such perceptual-cognition mechanisms of more ancient provenance (hence, at the lowest cognitive level) reveal more impressive competencies of deduction, inference, and physical intuition.

In focusing on just these two levels of cognition – the evolutionarily perfected and the recently improvised – we are apt to overlook what is perhaps the most remarkable fact about the extended phenotype of humankind. This is the fact that our internal cognitive capabilities (however incomplete, approximate, or heuristic they may still be) have nevertheless enabled us to develop the externalized and shareable representational systems needed to prove theorems of logic and mathematics – including those of Bayesian inference and of kinematic geometry, invoked here – that *do* accurately represent universal features of our world. This I denominate the third level of cognition.

Granted, this third level has emerged from the two lower levels. We use the recently developed, imperfect heuristics of the intermediate level of conscious deliberation to struggle toward our abstract formalisms of mathematics and physics. And our success in doing so is implicitly guided, I claim, by the more powerful and less fallible machinery of generalization, inductive inference, and mental transformation implemented deep within the lowest cognitive level.

Yet, the emergent third level is fundamentally different from both of its two lower-level progenitors. It differs from the lowest, evolutionarily perfected level in that it is itself neither genetically transmitted nor automatically performed. It differs from the intermediate, heuristic level in that (although it, too, may use paper and pencil and other external aids – most recently, computers) it achieves representations that reflect a far greater range of mathematical and scientific facts and principles, and does so with far greater accuracy. Indeed, it is by virtue of this third level that we can see that our second-level methods and representations are less than optimal and, in some cases, we can even specify the conditions under which they do or do not approximate optimality, and to what degree.

Granted, too, this third level of cognition, unlike either of the two lower levels, is not universally manifested – even in humans. The explicit knowledge that it has engendered, and the environment- and society-transforming technology to which it has led, have stemmed from the mental labors of those very few – out of the many billions of members of our species – who are the likes of Archimedes, Galileo, Fermat, Newton, Euler, Laplace, Gauss, Bayes, Darwin, Helmholtz, Maxwell, Chasles, Einstein, and Feynman (to recall names I have already invoked here).

Researchers who have adopted an evolutionary approach to cognition have tended to emphasize the continuity between humans and other animals and, also, the domain-specific nature of adaptations – whether human or animal. From my own evolutionary perspective, however, I have become impressed by the unique generality of the so-called "cognitive niche" into which the human line has emerged. Indeed, I have come to feel that the term "cognitive *niche*," with its connotation of confinement within a small region of the space of possible strategies, is not as appropriate as a term with a more open connotation, such as "cognitive *space*." In any case, with the advent of the human capacity for abstract thought and for language has come the terrestrially unprecedented possibility of standing back from immediate self-interest and considering, abstractly, the possible states or transformations afforded within some well-conceptualized realm. This, I suggest, has made possible two things of utmost significance for the destiny of



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

humankind: the discovery of the fundamental laws of nature and the formulation of the objective grounds of universal moral principles.

### *SR6.2. The cognitive grounds of science*

When I began looking into what discoverers of the fundamental laws of nature have said about how they arrived at their discoveries (Shepard 1978a and subsequent articles), I was struck by three things:

*First,* many such scientists have indicated that they achieved their insights through "thought experiments" using processes such as Einstein described as "visualizing . . . effects, consequences, possibilities" by means of "images which can be 'voluntarily' reproduced and combined" (Einstein, as quoted in Hadamard 1945, p. 142, and in Wertheimer 1945, p. 184). Such processes seem to be of just the nonverbal, "analog" type that proved to be, themselves, highly lawful in the simplified and purified forms in which my students and I studied them in our experiments on "mental images and their transformations" (Shepard & Cooper 1982; Shepard & Metzler 1971).

*Second,* many have remarked how, once their theoretical formulation had fallen into place, it appeared to have an almost mathematical necessity. Time and again they have echoed Bohr's exclamation, "What fools we've been! We [now] see that absolutely everything has to be exactly as it is" (quoted by J. A. Wheeler – in French & Kennedy 1985, p. 223). (Compare **BARLOW**'s remark, in his article here, "If I had been smart enough I would have predicted [what] Hubel and Wiesel . . . discovered.") In the same vein, Steven Weinberg has observed that each of the two culminating achievements of twentieth-century physics – general relativity and quantum mechanics – is so tightly constrained by its own self-consistency and symmetry that it appears virtually impossible to make any small adjustment without disrupting the whole structure (Weinberg 1992, e.g., pp. 86, 98, 102, 104, 211). Concerning the earlier Newtonian mechanics, one could ask why gravitational force decreases with distance raised precisely to the power −2.0 rather than some other, nearby value, such as −1.9 or −2.2. But in general relativity the unique value of −2 falls out as a necessary consequence of the geometry of a curved space-time manifold. It is presumably his awareness of such constraints that led Einstein to remark, "What really interests me is whether God had any choice in the creation of the world."

(Empiricists take note: Because general relativity and quantum mechanics are not consistent with each other, the discovery of a single self-consistent theory that subsumes both of these theories as special cases – much as they each subsume Newtonian mechanics as a special case – will be an achievement of the very first magnitude, and without collecting a single additional empirical datum.)

*Third,* theoretical physicists, in particular, have often claimed that among alternative theories, the one that possesses the greatest beauty, elegance, and symmetry is likely to provide the best approximation to the reality behind the phenomena we seek to explain. Again, to cite Weinberg: "I believe that the general acceptance of general relativity was due in large part to the attractions of the theory itself – in short, to its beauty." "And in any case, we would not accept any theory as final unless it were beautiful." Specifically with regard to symmetry, Weinberg wrote, further, that "it is principles of symmetry that dictate the *dramatis personae* of the drama we observe on the quantum stage." (Weinberg 1992, pp. 98, 165, 212, respectively.)

I turn now to a few simple illustrations of the power of these three guides to the discovery of fundamental laws of nature – thought experiments, self-consistency, and symmetry. I particularly focus on the role of symmetry, which, being mathematically defined as invariance under transformation, is fundamental in the theory of transformation groups (and, hence, directly related to the topic of my sect. SR2, on mental transformations). Indeed, invariance under transformation may be the basic reason for the perceptual system being so keenly attuned to symmetries (Shepard 1981b).

All three of Newton's laws of motion are directly derivable from thought experiments. For illustration here, I use only Newton's third law: "Every action has an equal and opposite reaction." Not only is this law (of the three) the most original with Newton (the other two having been anticipated, especially by Galileo), it also illustrates the power of symmetry most simply.

I imagine Newton imagining himself forming an arch between two identical boats in the middle of a lake, with his feet on the gunwale of one boat and his hands on the gunwale of the other. (I also imagine the two boats to be equally loaded with apples, though this particular cargo is not essential to the thought experiment.) From the symmetry of the imagined situation, Newton realizes that he could not push the second boat away from the first with his hands without simultaneously and equally pushing the first boat away from the second with his feet. (By an extension of the same reasoning, Sir Isaac, having fallen into the water between the two now separating boats and having lost his oars in the process, can conclude that by climbing back into either boat and vigorously hurling its cargo, apple by apple, in the boat's rearward direction, he could propel himself toward shore – just as spacecraft now propel themselves through empty space by ejecting the molecular products of combustion rearward at very high speed.)

For the second illustration, I imagine Galileo imagining himself atop the Leaning Tower at Pisa hefting three identical bricks. By the symmetry of invariance of identical objects under permutation, Galileo immediately knows that if he were to release the three bricks at the same instant, they must reach the ground simultaneously. By hypothesis, there is no reason for any one of these identical bricks to fall faster than any other. Now comes the critical step in the thought experiment. Galileo imagines that two of these three bricks are joined into a single larger brick by means of a length of string or a film of glue. Aha! thinks Galileo. The resulting twice-as-heavy brick would surely not, by the addition of a virtually weightless piece of string or dab of glue, now fall twice as fast as the third brick, alongside which these same two bricks, when unattached, would have dropped at the very same speed.

Here again a fundamental fact about the physical world is decisively reached by symmetry considerations through thought alone – without actually dropping any objects. And the fact thus achieved is not a trivial one. Later, as the fundamental principle of the equivalence of gravitational and inertial mass, it was destined to play a central role in Einstein's theory of general relativity. Yet, this same fact directly contradicted what had widely been believed for over 2000 years before Galileo, following Aristotle's proclamation that material bodies fall at rates proportional to their weight.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

For the third illustration, I imagine Archimedes imagining a horizontal beam resting on a knife-edge fulcrum, with any number of identical weights distributed along the beam in a way that does not cause the beam to tip down at one end or the other. Archimedes realizes that if any two weights on the beam are simultaneously moved to the same location exactly halfway between them, the beam must remain in balance. Drawing from his own cognitive "bag of tricks," the resourceful Archimedes need only imagine that those two weights (alone among all the other weights on the beam) are resting on a secondary beam (of negligible weight but sufficient rigidity) with *its* fulcrum (a) exactly halfway between those two weights and (b) resting on the primary beam. For, again by symmetry, as long as those two weights are equally distant from the fulcrum of the secondary beam, there is no reason for either end of that secondary beam to tip down, and this remains true no matter how close together those weights are moved. So, when they have both been moved to the location exactly above the fulcrum of the secondary beam, that secondary beam is no longer needed. The two weights can then be placed where they had previously communicated their combined weight to the main beam through that (assumed virtually weightless) secondary fulcrum.

Having thus justified the operation of moving any pair of separated weights to the midpoint between them on the main beam, Archimedes can now dispense with the imagined secondary beam altogether and simply imagine successively applying the procedure of moving to their mutual midpoint any two weights that still remain separated on the primary beam. With a little more thought he can verify that with continuation of this process, all the weights must converge to a single stack exactly above the fulcrum of the primary beam, and that this is the *centroid* of the original distribution of the weights. (He can also speed up the convergence by moving any two equal *stacks* of weights to their midpoint.) So here, through considerations of symmetry, Archimedes, using only mental manipulations and some mathematical calculations, can establish two things fundamental to physics: the law of the lever and the concept of the center of mass. And, just as in the case of Galileo, he can do this in the absence of any physical weights, without actually performing a single experiment or making a single measurement.

I have used Archimedes's law of the lever, Galileo's law of falling bodies, and Newton's law of action and reaction because they illustrate the role of symmetry in a particularly simple and direct way. But I could have given many other examples. How can this be? Physics is generally regarded as an *empirical* science. How can the truths of its fundamental laws emerge as necessary consequences of abstract mathematical principles – such as principles of symmetry? Contemplating the situations envisioned in such thought experiments, one feels the necessity of the conclusion in much the same way that one does when contemplating the diagram used to prove a purely mathematical fact such as, for example, the Pythagorean theorem for right triangles. "In a certain sense," as Einstein has written, it seems that "pure thought can grasp reality, as the ancients dreamed" (Einstein, in Schilpp 1949, p. 398).

Einstein, of course, is especially noted for his effective use of thought experiments – whether about light signals between observers in a (very fast!) moving train, or about the apparent path of a beam of light to an observer stationary on Earth versus one free-falling in an elevator. As Weinberg observed, "It is glaringly obvious that Einstein did not develop general relativity by poring over astronomical data" (Weinberg 1992, p. 104). Through such thought experiments, Einstein also brought to light fundamental symmetries – not only between gravitation and inertia, but also between electric and magnetic fields, between space and time, and (most famously) between matter and energy. Indeed, according to Noether's theorem – proved by the insufficiently celebrated mathematician Emmy Noether – every continuous symmetry implies a conservation law and, conversely, every conservation law entails a continuous symmetry. Group theory and its associated symmetries now pervade the foundations of theoretical physics, where, as Weinberg put it, "it is principles of symmetry that dictate the *dramatis personae* of the drama we observe on the quantum stage" (Weinberg 1992, p. 212).

Perhaps the single physical principle most relevant for my present purposes, however, is the universal principle of *least action* ("action," being the time integral of energy, has the units of a product of time and energy). Although intimations of this principle can be found in the writings of Aristotle, Hero of Alexandria, and Leonardo da Vinci, the first indications of its true potential emerged from Fermat's derivation of Snell's Law for the refraction of light from a least-time principle. (See my example of the least-time path for the lifeguard mentioned in my introductory sect. SR1 and in my sect. SR2 reply to **O'Brien & Opie**.) The principle was first formulated specifically as a least-action principle, however, by Maupertuis and (though apparently not independently) by Leibniz. It was then successively refined and generalized for classical physics by Euler, Lagrange, and Hamilton, and finally by Feynman for all of physics including quantum mechanics.

Already by 1886 Helmholtz had judged it "highly probable" that the principle of least action "is the universal law governing all processes in nature." Einstein's 1915 theory of gravitation, in which the trajectories of all unconstrained bodies are geodesics in curved space-time was later shown by Hilbert to comply, also, with the least-action principle. And in 1915 Planck, whose quantum of action, $h$, is the cornerstone of quantum mechanics, characterized the principle of least action as the "most comprehensive of all physical laws" and as one that "may claim to come nearest to that ideal final aim of theoretical research." Although the two major theories of twentieth century physics, general relativity and quantum physics, are not consistent with each other, the universal principle of least-action, being consistent with both, will presumably survive in any theory that subsumes both of those theories (such, possibly, as some version of string or D-brane theory).

According to Feynman's culminating "path-integral" formulation of least-action, all motions, including the quantum mechanical, are governed by a universal variational action-minimizing principle (Feynman & Hibbs 1965). It is significant for my purposes that this principle unites mechanical and teleological causation as two aspects of the same process. Light or a material particle, pursues, in mechanical (specifically *quantum*-mechanical) terms, all possible paths between its point of origin, A, and its point of later arrival, B. But the result appears as the traversal of a geodesic path (a locally straightest line or shortest path in a homogeneous medium) because all other paths cancel out through quantum-mechanical wave interference. The teleological result is that the motion turns out to have been, in the end, the one that minimized the total action. It is as if



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

the particle had at the beginning of its journey "sniffed out" (as Feynman has put it) the unique path that would minimize the action over the entire journey – even though the journey, as in the case of a photon from a distant quasar, may take billions of years. Feynman's path-integral formulation, sometimes referred to as "the democracy of histories," is fundamentally a symmetry law in that all possible histories are symmetrically treated, one might say, as "equal under the law."

There are several suggestive analogies between the universal least-action principle in physics and what I have been advocating for cognitive science. The resemblance of the least-action path to the shortest, simplest, least-time or least-energy paths that **Foster** and I have proposed for apparent motion and for mental transformation is clear. Also suggestive is the apparent similarity between the sum-over-paths derivation of the least-action principle in Feynman's quantum-mechanical formulation, on one hand, and the sum-over-hypotheses Bayesian derivation of cognitive principles of perception, learning, and generalization represented by my theory of generalization and by the extensions described here by **TENENBAUM & GRIFFITHS**. Finally, as I shall suggest in my concluding Section SR6.3, there may be significant parallels between the ways in which considerations of symmetry lead to the principle of least-action in physics and to a corresponding principle of best-action in ethics.

### *SR6.3. The cognitive grounds of ethics*

Having entered my 70s and the looming shadow of my own mortality, I have retired from teaching, turned my laboratory space over to younger researchers, and asked myself: What is the single unresolved theoretical issue of greatest human significance about which I might hope to gain some clarity of understanding during whatever time may remain to me? For me the issue was clear: Are moral principles discoverable that partake in any way of the universality and objectivity of the principles established by science? Or must we resign ourselves to the nihilistic view, common among scientists, of moral relativism – as if we are but fleeting flickers in the cosmic "screen saver" for an absent God?

As I began to struggle with this ancient and vexing issue, I was buoyed by the idea that the rather unorthodox view of science, cognition, and mind to which I had already arrived afforded a promising approach to the two seemingly most enduring and refractory problems of moral philosophy. The first is the general problem of the status of prescriptive norms in the objective world described by science – or, as one of the founders of Gestalt psychology put it, "the place of value in a world of facts" (Köhler 1938) or, as one might equally succinctly put it, the place of "ought" in a world that "is." The second is the notorious problem of providing an account of freewill that is compatible both with science and with the notion of moral responsibility. I turn, now, to the first of these two problems.

As one whose career has been spent laboring in the fields of science, I must emphasize at once that the problem of finding a universal objective basis for moral principles is *not* to be solved by the standard approach of empirical science. Arguments or evidence that natural selection can lead to species capable of behaving in some way that we regard as morally commendable (such as altruistically risking their own individual lives or well-being for the benefit of related members of their species), miss the point. Altruism evidently does arise through natural selection, but so do many behaviors that most of us find morally repugnant – including parasitism, slavery, mate abuse, infanticide, or prolongation of the painful death of prey (such as when a cat brings back the mouse, still living and suffering, to its kittens to provide them with an opportunity to sharpen their own hunting and killing skills). More to the point, to conclude that a behavior is good because it furthers the propagation of a species is to take for granted that the propagation of that species is good. We may believe that the perpetuation of our own species is to be desired, but this hardly provides an objective (that is, non-anthropocentric) justification for such a belief – or for any proposed moral law that might contribute to such perpetuation.

Nor is the problem of finding a universal, objective ground for ethics to be solved by any finding that there is some moral precept (such, perhaps, as the Golden Rule) to which all mature and cognitively functional human beings will assent. Such a finding could be explained as being derived, by generalization, from the altruism that evolved for the propagation of our own genes, as just considered. Or it could be explained as arising through a process of social evolution (based on "memes" more than on genes), in which societies that transmit such values to their children tend to flourish more than those that do not. Indeed, even if game-theoretic arguments showed how altruistic or golden-rule principles might be favored in cognitively advanced social beings wherever they evolve, the conclusion that such principles are therefore good presupposes that the furtherance of such intelligent social beings is good. It leaves unanswered the question of what, ultimately, is the objective justification for *that* presupposition.

Those skeptical of the possibility of establishing any ethical principles that are absolute (as opposed to being merely species- or culture-relative) point to the wide differences in what is regarded as ethical or unethical in different cultures, even within our own society – concerning such life-and-death issues as birth control, abortion, capital punishment, euthanasia, and animal rights. How can there be universal moral principles when such disagreements seem to defy resolution even within our own, scientifically advanced Western society?

My answer to this is that if objective, universal moral principles do in any sense exist, they are not to be discovered or justified through a consensus of public opinion – any more than mathematical theorems or the principles of Newtonian mechanics, relativity theory, quantum mechanics, or the theory of evolution are to be discovered or justified in that way. In fact, only a tiny fraction of the world's human population has anything approaching an adequate understanding of the principles and justifications for Newtonian mechanics or for Darwinian evolution – let alone, for relativity or quantum mechanics.

Opinions have differed about whether the universe is finite or infinite, whether the Earth is flat or curved, unique and stationary at the center of the universe or just one of countless planets, each orbiting about its star. Moreover, opinions still differ widely about the age of the Earth, about whether humans evolved through random variation and natural selection or through purposeful design, or about whether capital punishment deters criminals or gun control decreases the rate of homicide. But, despite the emotion, persecution, and even burning at the stake that the answers proposed to such questions have led to, they are all scientific


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

questions for which there *are* objectively correct answers – independent of how many people know, have known, or will know those answers. What, then, are the tenets of my particular view of science, cognition, and mind that give me hope that knowledge may be attainable concerning moral truths having the absolute, objective, and universal status of the truths discovered (or approximated) by science?

My first such tenet is that the concrete, material objects with which science is supposed to be exclusively concerned are not the only things that exist independently of our explicit knowledge of them. Some of these things are very abstract. They include logical relations of entailment and mutual consistency, and (in my rather Platonic view) the mathematical structures of number theory, geometry, and group theory. Other things, though equally abstract and nonmaterial, nevertheless have a physical existence. A striking example is an abstract relation that does exist between the not-yet-existent properties of two widely separated but quantum-mechanically "entangled" particles. Quantum mechanics requires that certain measurable properties, such as the spin orientations of such particles, have no determinate value until they are actually measured. At the same time, however, quantum mechanics requires that when and if those spin orientations are measured, they will be found to be opposite – if one is "spin up," the other will be "spin down," and vice versa. Likewise, some psychological things have a counterfactual, conditional physical existence, such as the orientation in which a match would occur if the appropriate stimulus were presented in that orientation during "mental rotation." (See my sect. SR2 reply to **KUBOVY & EPSTEIN**.)

Finally, though not abstract in the same sense, there are the consciously experienced phenomenal qualities (referred to in my reply to **Mausfeld** in the preceding sect. SR5 – see also, Shepard 1993). Such "qualia" (of which colors are the canonical example) are knowable only through one's own experience of them. No amount of scientific knowledge about the physical processes giving rise to them (whether waves or particles incident on a receptor surface, or ensuing electrochemical events in a brain) would enable a congenitally blind scientist, to mention the most frequently discussed case, to know what the *experience* of red or blue is like. Nor do we ourselves, have any idea of what the experience of color on a fourth dimension might be like for a person genetically endowed with a fourth spectral-sensitive class of retinal cones (a possibility acknowledged in my reply to **Decock & van Brakel** in sect. SR3).

My second tenet, which opens a little further the door to the possibility of truly normative (as opposed to merely descriptive) moral principles, is my epistemological realization that the just-mentioned phenomenal experience is in fact *all* that I directly know. All scientific knowledge lives only insofar as it is understood within the mind. The text, equations, and diagrams in any scientific article or treatise are but meaningless arrangements of molecules of ink on a molecular substrate of paper unless their meaning is comprehended within some mind. True, much of my own scientific knowledge has been gained, indirectly, through the qualia I have experienced in studying such papers or books (or in listening to the spoken words of other scientists). But I have come to accept such indirect knowledge (a) on the inductive grounds that where I have been able to test such indirect sources for myself, I have generally found them to be confirmed by my direct experience and, more frequently, (b) on the rational grounds that I have generally found information from such indirect sources to be consistent with the world picture that I have built up from all sources during my life. This is the epistemological approach to science articulated by the late Nobel Laureate for physics, P. W. Bridgeman (1940), and sometimes referred to as "methodological solipsism." Its principal relevance here is that it undermines the highly reductionistic materialism that for many scientists seems to preclude the possibility of absolute moral knowledge.

On one hand, the concrete material objects that familiarly populate our world are directly experienced only through their phenomenal qualia of color, size, shape, texture, heft, and so on. On the other, according to current scientific theories, at the smallest scale the so-called material particles out of which such objects are built are but particular modes of vibration of space-time – a rather abstract conception in itself.

My third and final tenet, closely related to the other two, is that moral laws may be discoverable by striving for self-consistency by means of thought experiments that make use of inner knowledge of abstract principles such, particularly, as those of symmetry. I suggest that this may be possible in much the same way that I argued earlier in section 6 that physical laws can be discovered by striving for an overall self-consistent system through thought experiments aided by a knowledge of abstract, nonphysical principles – particularly those of symmetry, as I illustrated for Archimedes's law of the lever, Galileo's law of falling bodies, and Newton's law of action and reaction.

Toward this end, I shall focus first on one roughly formulated candidate for a universal moral principle that the majority of humans, despite their disagreements on many particular issues, evidently will claim to endorse – even if their behavior often falls considerably short of full compliance. This is the already mentioned golden rule, which has been formulated in many different ways but for which the following statement may suffice at present: "Treat others as you would have them treat you."

In physics, as I noted earlier, the fundamental universal principle of least-action, which only reached its currently most general formulation with Feynman in the middle of the twentieth century, had a long evolutionary history that can be traced back through its first formal, if inadequate, statement by Maupertuis in 1744, to earlier less complete or formal statements going back to antiquity. Moreover the central idea has all along been implicit in the human desire, often and widely expressed on undertaking any task, to minimize one's time and energy.

Similarly, the golden rule has been traced back to Thales of Miletus before 600 BC, and has been stated, in some variant, as a fundamental tenet of virtually all of the world's major religions – including Confucian, Buddhist, Hindu, Jewish, Christian, and Islamic (Wattles 1996). Major philosophers have endorsed it as summarizing "immutable and eternal laws of nature" (Hobbes 1651, XV, p. 79) and as constituting "the ideal perfection of utilitarian morality" (J. S. Mill 1861/1863, p. 323). It is entailed by Kant's (1785/1996) categorical imperative; and it entails, in turn, Jefferson's (1776) "self-evident" truth that "all men are created equal, [with] unalienable rights . . . ."

Moreover, this long history, like that of the principle of least-action, has to some extent been an evolutionary one (again, see Wattles 1996). In some of the early formulations,



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

it appears that the "others" specified in the principle were only intended to be those belonging to one's own group – whether one's family, friends, tribe, state, race, or species. Moreover, the principle was sometimes justified only pragmatically, through the benefit it confers on one's self by the fact that others generally tend to treat one as one has treated them. But later versions tended to recognize that the principle was morally, rationally, or at least psychologically binding, independent of self-interest. Thus John Wesley said that it "commends itself . . . to every man's conscience and understanding; insomuch that no man can knowingly offend against it, without carrying his condemnation in his own breast" (Sermon XXV, 1742; in Sugden 1921, p. 529).

But no matter how widely the golden rule has been put forward or how universally it has been endorsed by humankind, the question can still be raised whether such a rule is in fact morally and rationally binding. Many – perhaps the majority – of scientifically oriented contemporaries take the position of moral relativism in which every being is considered to act solely out of self-interest, or – as may be further specified by evolutionarily oriented scientists – out of self-interest that, in large part, reflects the "self-interests" of one's genes. Anyway, as I have stated in connection with my own psychological research, the empirical finding of some uniformity (such as the exponential shape of the generalization function) is not in itself sufficient to establish that uniformity as a universal psychological law. For that, we need, in addition, a rational argument as to *why* the uniformity found in humans or other terrestrial animals should be approximated in all cognitively advanced beings.

Thought experiments do appear to have played just as significant a role in the development of proposed moral principles as in the development of physical laws. In support of his categorical imperative, "Act only on that maxim through which you can at the same time will that it should become a universal law," Kant (1785/1996) presented thought experiments concerning, for example, truth-telling and promise-keeping. As he noted, persons who lie or break their promises undermine the effectiveness of their own future lies or promises. This might be seen as basing moral law, once again, on self-interest; but Kant clearly regarded it as rationally binding "not merely for men, but for *all rational beings as such.*" More directly relevant for me is the principle proposed in the highly influential *Theory of Justice* of John Rawls (1971), which I paraphrase as follows: "In framing rules to govern a just society, proceed as if you cannot know which role you will be assigned in that society." Without this "veil of ignorance," Rawls noted, framers cannot help but favor rules that will ensure the preservation of advantages of wealth, property, position, or power that they already posses at the expense of those less fortunate. This, too, is a thought experiment. There is, of course, the practical question of how one could ever persuade the currently more well-situated to agree to submit themselves to the veil-of-ignorance ground rules. But what the thought experiment most directly provides is conceptual clarification, not a practical basis for social or political action.

My proposed grounding of ethics has four essential parts, which must now be explicitly recognized and explicated. Just as I have claimed for scientific principles, the discovery and the understanding of moral principles occur only in the mind. From the epistemological standpoint, some parts of this grounding are accordingly best stated in the first-person voice.

*First,* through my own direct experience – just as certainly as I know any empirical fact (such as, that the sun will rise tomorrow, that the ground will support my next step, or that if I release this heavy brick it will drop) – I know that pain, suffering, debility, and restriction on my freedom of action and access to knowledge are bad. Correspondingly, I know that the avoidance or cessation of such pain, debility, or restriction is good. (I may, of course, freely choose to suffer pain – as in a dentist's chair – in the expectation that this will reduce the likelihood of a subsequent greater pain or debility.)

*Second,* – just as firmly as I accept that behind the often transitory phenomenal appearances that are my experiences of rocks, trees, animals, and other people, there are enduring physical objects that exist independently of my direct phenomenal experience of them, – I also accept that behind the phenomenal appearance I experience of another person, there is another mind that exists independently of my interactions with it, which also has phenomenal experiences, including pains and pleasures, much as I do. It is tautological that what I directly experience is only my own phenomenal experience, and not that of the independently existing physical bodies or other minds, as such. (Indeed, for a short period in my youth I became preoccupied with such possibilities as that I am just a brain in a vat or, more fundamentally, that my immediate experience is *all* that exists and that there is no external world and hence no physical brain, or no past, or no other mind!) I now feel, however, that the inferential leap I naturally make to belief in the existence both of external bodies and of other minds is rationally justified from the orderliness of the phenomena I do directly experience. This pereceptual inference, which is so automatic and compelling, might even be regarded as a kind of Gibsonian "pick up" of "higher-order" aspects of my phenomenal experience.

*Third,* having thus rejected solipsism and accepted the existence of other experiencing minds, my deeply internalized intuitive grasp of a principle of symmetry compels me to embrace some form of the golden rule and of abstract moral principles such as that of justice. Indeed, I suggest that the single most fundamental intuition behind the golden rule, Bentham's and Mills's utilitarianism, Kant's universalizability, Jefferson's Bill of Rights, and Rawls's theory of justice – despite their obvious differences – is that of symmetry. Specifically, in the spirit of the abstract group-theoretic conceptualization I have been advocating, it is the symmetry of invariance under permutation of the members of an equivalence class.

*Fourth,* (and finally) – just as surely as I know any empirical fact – I also know that I am free to decide what I should do (even under the circumstances in which I am externally constrained from overtly acting on that decision). Moreover, such freewill is essential for the existence of an objective morality because, without it, I could not be morally responsible for any decision (or for any overt action I take based on that decision). But how is this essential freedom of the will to be reconciled with the fundamental tenet of science (and, indeed, of rationality itself) that any event is either determined in advance (e.g., by prior physical causes) or is merely random (as allowed, e.g., by quantum theory)?

Still more challenging, how is the idea of moral responsibility to be reconciled with the idea that a (tenselesss) assertion about a particular event occuring at a particular time


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

and place, must be either true or false for all time, and hence far in advance of the specified time – even though we cannot, of course, *know* whether the assertion is true or false until *after* that time? This question arises whether the event is determined by prior physical or nonphysical causes, has no prior causes, or is an act of our own freewill. (Notice that the alternative – namely, that the statement about an event's occurence has no truth value *until* the time specified for its occurrence – raises problems that no proposal of a temporal logic seems to have adequately addressed. For example, the question of the *speed* at which the truth value is supposed to propagate out through physical space-time from the specified time and place of that physical event.)

For more than 50 years, since I first began worrying about the problem of free will, I had seen no way in which freewill (any more than qualia) could be reconciled with the physicists' conception of the world. Then, in just the last few years, the view of rationality to which I have arrived seemed to me to provide an unexpectedly simple solution to this problem.

As I have mentioned, I have come to believe that in the case of terrestrial evolution, a new cognitive capacity has emerged in the human line. This is the capacity for abstract thought and for language. (Elizabeth Spelke, personal communication 2000, reports coming to a very similar conclusion, though she attributes relatively more importance to the emergence of language.) This includes, in particular, the capacity to "stand back" from one's immediate biological needs, self-interests, and pre-wired domain-specific strategies. One is thus enabled to consider the set of possibilities afforded by some well-conceptualized system, to recognize the abstract symmetries in such a set of possibilities, and to formulate general strategies for attaining explicitly formulated goals within such a system.

Within a well-defined set (A) of alternative objects or actions, a cognitive agent represents the subset (B) of alternatives that are equally suited to achieving a specified goal and that are better than all the other alternatives within the set A for that purpose. The agent is free to choose any item in the set A, that is, the agent feels no constraint in contemplating the initiation of any of the alternative choices. Being free in this sense, the agent chooses an alternative from the subset B, within which the agent has calculated that all alternatives are more likely to lead to the adopted goal than any alternative outside that set. Such a choice is not uncaused or undetermined, but the cause or determination is based on an abstract logical assessment or calculation and is, in a sense, a final cause as well as a proximal one. Within the subset B, the agent's calculation has yielded no preferences, so the agent can choose any alternative. In terms of causal mechanism, the agent may use a selection module that is either truly random (as through a quantum mechanical process), or a module that merely appears random to the agent because the underlying micromechanical process (which may be chaotic or thermal) determining the outcome is not accessible to the agent's own cognition. So the choice is subjectively free but leads deterministically to an alternative in the preferred subset (B). In the extraordinary event that the agent's overt action corresponds to an alternative outside the set B, the agent immediately knows that his or her action was not free and, hence, that he or she is not morally responsible for the action. In the usual situation in which the agent's action corresponds to an alternative within the set B (typically, the particular alternative chosen within the set B), the micromechanical process that led to the selection within that equivalence class is cognitively and morally irrelevant to the agent.

In short, all the worrying about how one chooses freely among alternatives has missed the essential point. On one hand, if there is no reason to choose one alternative rather than another, then there is no moral issue, and the decision can indeed be made by a process that, though probably causally determined, is not determined by a constraint of which the agent is aware and, for this reason, seems unconstrained – that is, free. If, on the other hand, the agent has a moral reason to choose one alternative rather than another, then that reason is also a cause. But because the causal determination is the agent's own computation of the preferred, just, or moral choice, the agent experiences no constraint opposed to his or her choice and, in this sense, is again free.

Many philosophers hold that to claim that one chose freely is to claim that one *could have* made a different choice. They might say that I have not established the existence of true freedom. But, they need to state just what sort of freedom they seek. Surely, a different choice would not have been *my* choice, if that choice was determined by a wholly random (e.g., quantum mechanical) event. It would only have been *my* choice, if *I* had made it. To be counted as a moral choice, moreover, I must have made it for *reasons.* What my freedom really consists in is not that the choice is undetermined but that it is determined by *my* reasons. When I say that I should have – and by implication could have – chosen differently in a particular circumstance is that, given what I now know or believe, I have reasons for choosing differently when and if that circumstance arises again.

This is not to say that I believe there are no unresolved issues facing my proposal for finding a grounding for ethics. There are several. I shall only take time to mention here the one that I believe poses the most formidable challenge to cognitive/social science. This is the problem of specifying which "others" should constitute the equivalence class for which a moral principle such as the golden rule should be invariant under permutation. Or, if we are to admit degrees of membership into the relevant class, what are the appropriate factors for determining such degrees of membership or equivalence? Are they factors having to do with attributed degree of sentience or susceptibility to feelings of pain or pleasure (as hedonic utilitarianism seems to suppose)? Or are they factors having to do with attributed degree of rationality (as Kant advocated)? I suggest that the former is most relevant to determining how we should treat others, and this is why I oppose the mistreatment of animals that seem to be just as affected by pain and suffering as we are (Shepard 1993). I suggest that the latter factors are most relevant to determining who should be held morally accountable for their actions, and this is why we generally do not hold very young children, mentally incompetent, psychotic, or demented human adults, or nonhuman animals morally responsible for their actions.

The division is not, however, sharp. The rationality consideration is relevant at least to some degree and under some circumstances to our treatment of the unborn fetus and even of such lowly forms of life as bacteria. The human fetus has the potential of *developing* into a rational adult, and even bacteria have the potential of eventually *evolving* into rational beings (if, for example, all other terrestrial life were to be destroyed by some cataclysm). And the sentience consideration is relevant to the morally responsible status of a



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

rational agent, because a capacity for purely cold, rational calculation and logical deduction (essential for the last two of the four parts of the argument stated above) do not alone provide grounds for morality. We need, in addition, the experience of feeling and the awareness of the feelings of others (as stated in the first two of those four parts above).

One thing that seems to be required, then, is some metric of similarity enabling inductive generalization from our own experience to the attribution of feelings and rationality to others. (Although I have not taken time to go into it here, I have been continuing to extend my work on generalization to the inductive grounds of science – Shepard 1997.) With regard to other terrestrial life, the generalized attribution of feeling should presumably be much wider than the generalized attribution of rationality. With regard to possible extraterrestrial life as well as artificial or robotic "life," however, a metric of similarity, while still important, may not be sufficient. Conceivably, there may exist, or come to exist, beings that are vastly more advanced than we in their rational and moral sensibilities, and such beings may not resemble us as closely as do chimpanzees or, indeed, as do cetaceans or cephalopods. (Certainly, above us on the scale of rational and moral development, there is plenty of room for such beings – as the widely accepted notion of God attests.) Perhaps, in addition to a metric of similarity, we need a vector representation that permits extrapolation in a direction of rational and moral sensibility exceeding our own.

The full resolution of these and other difficulties will require major conceptual and formal advances going well beyond the simple notion of the symmetry of invariance under permutation of elements within a single, sharply bounded set of "others" sketched here. Nevertheless, I am encouraged by the similarities between the evolutionary development of the symmetry principles of least action in science and of "best action" in ethics. (A similarity that is further recognized in the Taoist idea that best action is, in the sense of *Wu Wei*, least action.) I hope, anyway, that there may be some, among those who read this, who will begin to see, as I now do, this possibility: That human action need not always be determined by the self-interest of ourselves or of our genes; nor primarily by the reasoning that the action will maximize our happiness. Humans, I suggest, are capable of action determined, rather, by what is morally right or just.

ACKNOWLEDGMENTS

I am grateful to Rainer Mausfeld and Dieter Heyer for organizing the Zentrum für interdisciplinäre Forschung (ZiF) program, Bielefeld University, devoted to an examination of the possibility that has motivated my work in recent years – namely, the possibility that universal mental principles have arisen as accommodations to universal features of the world. Additionally, I am indebted to Laurence Maloney for his role in editing this special issue of *Behavioral and Brain Sciences*, which grew out of those ZiF meetings. Because I was unable to participate in those meetings, I have welcomed this opportunity to clarify my views and to address those of the commentators. I have also benefited from discussions with a number of colleagues, including Piet Hut and Arthur Zajonc in the physical sciences and Joshua Tenenbaum and Elizabeth Spelke in psychology. Finally, I thank the National Science Foundation, which supported my research through successive research grants for more than 30 academic years, first at Harvard and then at Stanford.

## Barlow's Response

## The role of statistics in perception

Horace Barlow

*Physiological Laboratory, Cambridge CB2 3EG, United Kingdom.*
**hbb10@cam.ac.uk** **www.physiol.cam.ac.uk/staff/barlow**

**Abstract:** Hoffman is worried that perception itself leaves no time for the computation and compilation of statistics, but this has never been proposed. It is the underlying mechanisms that are thought to have evolved in response to the statistics of sensory stimulation, and which are capable of adjusting their parameters in response to changes in these statistics.

**Hoffman** epitomizes my view of movement perception ". . . as resulting from cortical computation of the difference quotient," but that does not correctly summarize my view. Some facts are known about how cortical neurons achieve motion selectivity, and also about the mechanisms by which their patterns of activation are analysed. But a great many of the principles and details are still matters of speculation, as is the whole question of how the objective activities of neurons are related to the subjective experiences of perception.

I am much more perturbed by **Hoffman**'s failure to understand the role that I (and many others) think statistical regularities play in perception. He says: "As to *statistical regularity,* perceptions occur in real neurobiological time, without the time-consuming computations required for estimation of statistical parameters" (his emphasis). Computing statistics on the fly from single perceptual experiences makes no sense. I have never suggested it and I don't believe anyone else has either. The idea is that the statistical regularities of natural images determine what forms of representation are advantageous and what forms are disadvantageous. The suggestion is that the brain employs advantageous forms of representation that have been selected through ordinary evolutionary mechanisms, and that physiological adjustments occurring on very much shorter time scales add further advantages. Sometimes, as with the adjustments of light and dark adaptation that occur in response to changes of mean visual luminance, it is obvious that both genetically determined mechanisms and physiological adaptation are involved. It seems likely that this is true for many, if not all, of the mechanisms that exploit regularities.

I am also puzzled by **Hoffman**'s distinction between "regularities" and "statistical regularities." High spatial frequencies are attenuated in natural images, but to establish that such attenuation is very generally present requires statistical evidence. And this must be true for any forms of departure from randomness – that is, for any regularity. Is there is any such thing as a regularity that has no effect on the statistics of sensory stimuli, or that can be known to exist in such stimuli without statistical evidence?

I love **Hoffman**'s broad conceptual scope and his wonderful extra-terrestial reach, but what we really need is down-to-earth knowledge and understanding of how the brain exploits environmental regularities.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

## Hecht's Response

## Universal internalization or pluralistic micro-theories?

Heiko Hecht

*Man-Vehicle Lab, Massachusetts Institute of Technology, Cambridge, MA 02139.* **hecht@mit.edu** **mvl.mit.edu/AG/Heiko/**

**Abstract:** In my response I revisit the question whether internalization should be conceived as representation or as instantiation. Shepard's ingenuity lies partly in allowing both interpretations. The down side of this facile generality of internalization is its immunity to falsification. I describe evidence from 3-D apparent motion studies that speak against geodesic paths in cases of underspecified percepts. I further reflect on the applicability of internalization to normal, well-specified perception, on the superiority of Gestalt principles, as well as on the evolutionary and developmental implications of the concept. The commentaries to the target article reveal an astonishing lack of agreement. This not only indicates that a satisfactory unifying theory explaining perception in the face of poorly specified stimuli does not exist. It also suggests that for the time being we have to be pluralistic and should treat internalization as a source of inspiration rather than as an irrefutable theory.

I would like to thank the large number of commentators who responded to my paper for their thought-provoking ideas. I take the surprising scope of the commentaries, which range from enthusiastic support of my challenge of Roger **SHEPARD**'s concept of internalization to its complete dismissal, as an indication that I undertook the right step. The concept has been pushed along on its non-linear path. There seems to be some disagreement as to how to define internalization. The empirical evidence is not the subject of disagreement but rather its implications for SHEPARD's theory. I shall first argue that internalization has to be understood as representation and as instantiation. I will rephrase my falsificationist argument where I think it may not have been sufficiently clear, in particular as far as my method is concerned. My method was to challenge the concept by exhausting possible interpretations of internalization, as well as by playing with the alternate concept of externalization. Rather than addressing individual comments, I have identified a number of topics, each of which comprises several comments.

### HR1. Internalization revisited

The responses reveal that different commentators have very different notions of internalization. Compare, for instance, internalization as a metaphor, as Michael **KUBOVY** and William **EPSTEIN** treat it, with the idea of internalization as neural space, as suggested by Shimon **Edelman**, and with the outright rejection of the concept by Andrew **Wilson** and Geoffrey **Bingham**. When I introduced a taxonomy of what could be internalized, basing it on the criterion of resolution, I had assumed some consensus on what constitutes internalization. The variability of the comments indicates that such a consensus has not been reached. This may be so because the concept is more or less a non-statement (see, e.g., Balzer et al. 1987). It does not contain testable substance and it is sufficiently vague to get away with it. Consequently, arguments about the essence of internalization are even more misguided than attempts to find experimental evidence for or against the concept. It matters little how we define internalization as long as we do not define it so narrowly that it immediately becomes implausible. Let us revisit the concept at its definition stage. I believe that two related concepts lie at the heart of our joint attempt to grasp internalization: instantiation and representation.

**KUBOVY & EPSTEIN** analyze the notion of internalization and conclude: "For the cognitive constructivist, the perceptual system *follows* rules; for the computationalists, the system *instantiates* them" (sect. 1.1, para. 7, emphasis theirs). They continue to juxtapose these two varieties of internalization with the Gibsonian position that, supposedly, is incompatible with internalization. This confines internalization to a lesser concept, more limited in scope than is suggested by **SHEPARD**. However, if the sacrifice of limited scope is not rewarded by an increase in explanatory power, this may be the worst of all cases, and we may indeed find that we can live without the thus confined concept. Instead, I believe that Shepard deliberately used internalization as both representation and instantiation without excluding ecological concerns. By doing so, he has traded diminishing explanatory power for universality. The universality lies in the accommodation of both representation and instantiation as ways of internalization.

### HR2. Representation or instantiation?

It is all too obvious that the perceiver does not explicitly represent the knowledge about the physical world when she experiences a percept of smooth apparent motion. Perception happens; it is an automatic process. Accordingly, representationalists do not claim that environmental structure is explicitly represented inside the mind in the way that memories of past events are. Rather, world regularities appear in the mind (or the brain) in some indirect manner. Gerard **O'Brien** and Jon **Opie** draw a similar conclusion when they offer an additional form of representation outside the realm of perception. The fact that the solar system represents (or instantiates) Newton's laws is an example for this variety of representation. Thus, we need a tripartite distinction. The first, instantiation in the world, is not very exciting for our discussion. Presumably all laws of physics are instantiated in all matter to which they apply at all. The second is implicit representation or instantiation as the outcome of a perceptual process, and the third is explicit representation in a cognizant system.

If the visual system uses acquired knowledge about world regularities to disambiguate questionable stimuli, then how could representation differ from "mere" instantiation? I see, as the first of two possibilities, that several representations are available and a choice between them is somehow made, while the system can only be one instantiation. The second possible distinction is that representations can be learned and reprogrammed, whereas instantiations are innate or hard-wired. For instance, if different regularities are obeyed at different times, the makeup of the system (instantiation) can no longer be held solely responsible.

The term instantiation has been used in many different ways in our discussion, but I do not see how it differs from representation other than along the just mentioned dimension of reprogrammability. The computational use of instantiation suggested by **KUBOVY & EPSTEIN** seems to make an unrelated distinction, namely whether or not we like to treat the visual system as an agent, which we often do, not only by postulating a homunculus who looks at the



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

retinal image and then draws conclusions unbeknownst to the perceiver. For, nobody seems to entertain the view that the visual system explicitly follows rules. Helmholtz's (1894) inferences are unconscious and Rock's (1983) "logic of perception" is implicit. Explicit representation seems to be trivially false. And because it is trivially false we should have no trouble agreeing to the view that internalization must refer to implicit knowledge or knowledge that merely arises by virtue of the hardware. Gibson and his followers are in agreement with this broader sense of instantiation. For example, the notion of perception as a smart device (Runeson 1977) states that the makeup of the visual system lets it behave as if it followed complex rules although it is behaving rather simply.

Consequently, implicit representation and instantiation turn out to be the same. **SHEPARD** wisely uses both in the context of internalization and he equally wisely does not take sides on the issue of reprogrammability. For instance, he states that "genes . . . have internalized . . . pervasive and enduring facts about the world" (target article, Introduction, p. 581). This would suggest that he conceives of internalization as innate instantiation. But he likewise talks about learning of regularities and representation. This openness has the additional advantage of avoiding the unresolved debate of direct perception vs. representation. If we follow SHEPARD's broad definition of internalization we no longer have to decide whether perception and cognition are qualitatively different, as David **Schwartz** suggests in his commentary, or if they are similar. Giving internalization a superior status might have freed it from many old debates. This great advantage is unfortunately linked to the disadvantage of losing ever more definition and explanatory power. The compromise of leaving internalization vague amounts to reducing its content to the statement that the visual system cannot but do what it is doing. How fruitful can this be?

It is interesting to speculate if the constructivist notion of a self-organizing percept, such as **Frank et al.** imply, has to be treated as a third way of internalization. In the context of kinematic geometry they conclude that "there is no need for any internalization of the screw displacement mode itself." They view apparent motion paths as the emergent outcomes of a neural process of self-organization. This merely seems to introduce an intermediate concept. Now the question is no longer "what has been internalized?" but "what constrains the self-organization in such ways that we perceive curves instead of differently shaped paths?" It might be in place here to remind us of the initial problem, which was the question of how the visual system deals with situations that are poorly defined, where the stimulus is degraded, intermittent as in apparent motion, or completely absent as in imagery.

## HR3. My falsificationalist view

The statement that a regularity R has been internalized is usually taken to mean that some people exhibit behavior in agreement with R some of the time. In the case of the circadian rhythm the evidence is very strong. Most groups that have been tested in complete isolation adopted a cycle of activity and rest that came very close to 24 hours when they were left to their own devices and sheltered from natural daylight, clocks, communication, and so on (Czeisler et al. 1999). Here a moderate falsificationist (compare Popper 1935) would acknowledge that the hypotheses of an internal clock has withstood attempts to prove it wrong. However, other conceivable examples for internalized regularities do not fare as well. Typically the evidence is very mixed. In apparent motion geodesic paths are not seen in 3D (see below), horizontality is only sometimes applied to the perceived stimulus, and so forth. Should we take only the cases that work as evidence for internalization while we conveniently ignore the abundant counter-examples? If we do so and drop the requirement that the behavior has to be consistent we have indeed found evidence for internalization. However, this reduces the original hypothesis to an existence statement. Internalization exists. We can rest the case as soon as we have found one instance of the hypothesized behavior. I find this unsatisfactory. By choosing examples that range from very broad Bayesian probabilities to very literal regularities, such as the regularity of gravitational acceleration, every single candidate for an internalized regularity lives up to the weak existence statement. Some evidence can be found for almost every single regularity that comes to mind. I believe that we should demand more from a theory. Counter-evidence should be able to speak against the theory. For this to be possible, we need to demand that the internalization hypothesis be phrased in such a way that it keeps some generality.

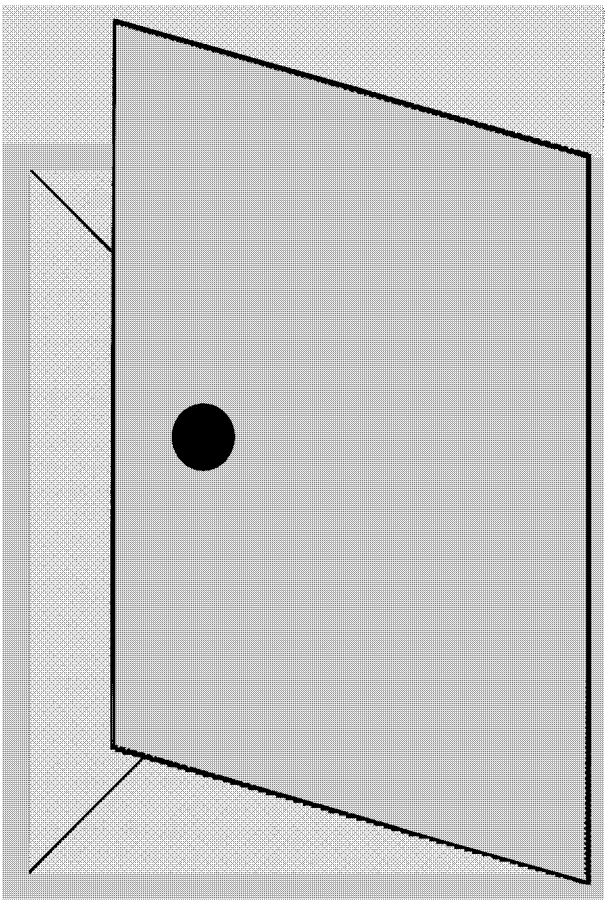


Figure HR1. Do internalized assumptions about object regularity fail to disambiguate the percept?

As long as internalization comprises both learning and instantiation, ontogenetic and phylogenetic knowledge, the concept is immune against all criticism. It may not


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

even be a metaphor. Does internalization of regularities at least rule out any class of percepts? Can we deduce from it that inherently ambiguous stimuli always get resolved one way or the other? Even this is questionable. While a Necker cube is bi-stable, that is, either one or the other interpretation is seen but never both at the same time, other such figures remain perceptually ambiguous. An interesting example is the ambiguous Figure above (see also http://www.illusionworks.com/). It looks funny. Something seems wrong, but the percept does not flip. Upon closer inspection we notice that the left edge of the dark panel is and is not at the same depth. Some internalized assumption about evenness suggests that the panel is rectangular in 3D while occlusion relations suggest that it is skewed. Here the visual system does not make up its mind. Does this mean nothing has been internalized? No. Does it mean that several things have been internalized? Maybe. The fact is that the unresolved stimulus can also not speak against internalization because the concept is immunized.

## HR4. Kinematic geometry

Since **SHEPARD**'s example of perceived apparent motion paths has received a lot of attention, I would like to add a few comments. While we are all in agreement that typically some degree of curvature is perceived when two differently oriented objects are presented in alternation, the distinction between 2D paths and 3D paths seems critical for Shepard's argument. Chasles' theorem clearly predicts circular arcs in 3D. Shepard takes the theorem and does what a good falsificationist should do. He adds specificity to the notion of internalization and operationalizes it for apparent motion in depth as well as in the plane. Hence the testable hypothesis is that unconstrained apparent motion should be perceived along paths that correspond to geodesics. That is, if the stimulus is 2D then a circular arc should be seen. If on the other hand the stimulus is 3D, geodesics that follow helical motions in 3D should be perceived. Hecht & Proffitt (1991) have put exactly this prediction to an empirical test by using a window technique.

Neither **SHEPARD** nor Dejan **TODOROVIČ** seem to have appreciated a fundamental disagreement of our data with the theory. Our Experiment 3 presented displays to observers that showed perspective renditions of dominoes in apparent motion. The stimuli were designed such that all orientation differences could be resolved corresponding to a single rotation in 3D. However, the 3D solution sometimes was considerably steeper and sometimes considerably shallower than the 2D solution that treated the objects as 2D blobs on the screen. Observers first had to adjust a tiltable plane such that it coincided with the perceived plane of the domino's motion. This was done with good accuracy indicating that the domino was indeed perceived to move on the plane in depth that was specified by its orientation. Figure HR2 shows the setup. It also depicts the 2D and the 3D solutions for this particular case, which were of course not visible to the subject. No matter how extreme the orientations, the window probe was practically never set beyond the 2D circle (labeled 100). That is, even though the motion was perceived in depth it fell far short of the circular 3D arc predicted by Chasles's theorem. This can only be interpreted as strong evidence against the internalization of 3D geodesics. The mere fact that the objects follow a curved path is insufficient evidence for the particular operationalization of kinematic geometry.

This piece of evidence against kinematic geometry seems symptomatic for the state of the concept of internalization. In its general format it is a non-statement. It has little or no testable content. The succinct operationalization offered by **SHEPARD** himself is testable and the evidence speaks against it. Now the question is, does it make sense to salvage this particular operationalization by claiming that geodesics are internalized but very sloppily so? This would only be reasonable if we had other examples where the percept does follow the predicted 3D path. Until we find such evidence we should regard it as falsified.

Marco **Bertamini** and **TODOROVIČ** rightfully point out the inadequacy of kinematic geometry as an explanation for how apparent motion paths are perceived, and even as a description of apparent motion paths. Interestingly, both – Bertamini more so than Todorovič – agree that the internalization of laws of physics is fundamentally different from the internalization of rules of geometry. It is not clear to me why they want to make this distinction and thereby go beyond **SHEPARD**'s claim that geometry is more deeply internalized than physics. I take Shepard to merely put forth a generalization argument stating that what is internalized has to be at an abstract level more akin to geometry. Bertamini and Todorovič ascribe two different internalization mechanisms to Shepard. Such a dichotomy, however, appears to be a category error: a geometric principle, once obeyed, is indistinguishable from the internalization of a physical law that predicts the same outcome. Roger **SHEPARD**'s phrase that "geometry is more deeply internalized than physics" (target article, sect. 1.5) is unfortunate because an internalization necessarily has to be some abstraction. Just as a law of physics is an abstraction as soon as it is isolated and used for prediction. Take an object falling from a resting position. Its motion is usually described sufficiently by the rule that it falls straight down. However, this is an abstraction because an indefinite number of other factors influence the object's motion. Thus, any internalization

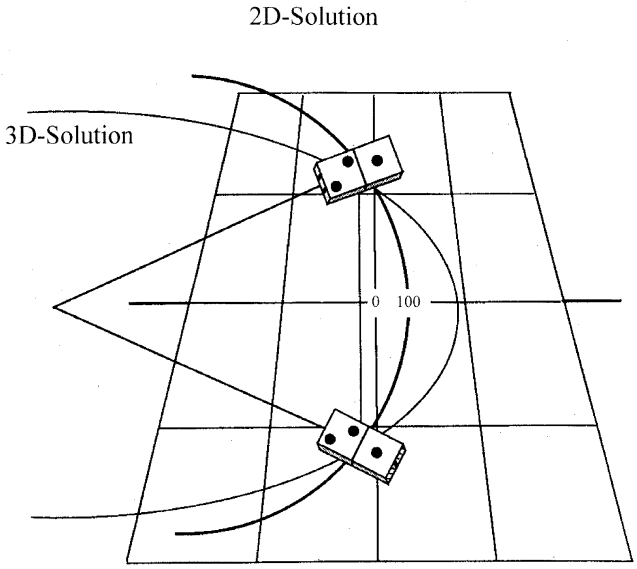



Figure HR2. Oriented objects in apparent motion that should be perceived according to the 3D solution but are not.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

necessarily has to extract an idealized or abstracted regularity. Or put differently, the full dynamics of a physical reality could not possibly be internalized. The distinction between physics and geometry can consequently only have a much subtler meaning along the lines that only very abstract rules become internalized.

Thus, stating that physical principles provide a better explanation than kinematic geometry is a result of taking the discrepancy between physics and geometry not for what it is. Internalization is necessarily approximative. Take, for instance, the fact that air resistance often distorts motion trajectories. Suppose that observers prefer squished curves that come closer to what happens with air resistance as opposed to paths in a vacuum. One could claim that effects of air have been internalized, presupposing a complex understanding of that force, or one could claim that squished curves are internalized. The outcome is not qualitatively different. The only thing that changes between these two "internalizations" is the *specificity* of the knowledge about an aspect of the world.

Incidentally, **Bertamini**'s torque explanation of the stimulus presented in his Fig. 1b seems a little far-fetched. The preference of the bottom path could just as well be attributed to differences in air resistance. We would then not need assumptions about a uniform mass distribution of the rectangular objects as are required for the moment of inertia (torque) explanation. This demonstrates that **SHEPARD**'s endeavor was justified. If the visual system has limited resources it should internalize a few general commonalities rather than a large number of specific rules.

David **Foster**'s remark that apparent motion (AM) may be an "internalization not of the ways in which objects move freely in space but of the ways in which observers manipulate or interact with them," basically takes my suggestion of externalization and applies it to the realm of apparent motion. What else could the internalization of ways to manipulate objects be? Since these ways are certainly not located in a world external to the observer, it seems awkward at best to call it internalization. Thus, we seem to differ merely at the level of semantics. I hold that we do not need a representation of the constraints on our effectors. Rather, these constraints seem to affect our representation of the world. But semantic issues aside, I have serious doubts whether this idea is as testable as Foster wishes it to be. Our preferred ways of manipulating objects may not be independent of the way we see them move.

This commonality of perception and action has been discussed in the domain of motor learning. The theoretical framework of common coding (Prinz 1992; 1997) posits that the final stages of perception and the initial stages of action control share a domain of coding where planned actions are represented in the same format as are perceived events. One of the implications of this approach is that, under appropriate conditions, perceived environmental events can induce certain actions by way of similarity or feature overlap. If perception and action share the same codes, it must also be expected that changes in these codes that are due to motor learning are reflected in corresponding changes in perceptual skills. Hence, the principles governing our imagery and unsupported perceptions may be just as much influenced by our motor constraints as by external events. The makeup of our effectors constrains what we are able to perceive and learn just as much as external regularities do. Our body dynamics also set the conditions for what principles may be applied in situations of ambiguity. Moreover, shared representations would necessarily have to be in approximate agreement with external and internal events.

## HR5. Evolution

The comments by Ken **Cheng** and Adolf **Heschl** shed an interesting light on my attempt to find regularities at very abstract as well as at very concrete levels. The example of birds extracting the position of the north star shows that very specific knowledge can be exploited. This contradicts the above mentioned weak claim that only abstract principles are internalized. Thus, at least from an evolutionary point of view specific regularities, such as water remaining invariably horizontal, are not too fine-grained to be in principle internalized. Heschl's argument that every existent perceptual-cognitive mechanism must have originated by an evolutionary process of adaptive internalization (i.e., eliminating negative and preferring positive outcomes) supports this. It also underscores my suspicion that internalization is a non-statement. Given that the theory of evolution is accepted, positing internalization does not add anything.

John **Pickering** shows that calling the process of internalization "evolutionary" does not mean stable instantiation. Lamarckian change even at short time scales suggests that some constraints that we find today may have been acquired very recently. This negates any possible differences between representation and instantiation, and it shows that ad hoc internalizations are compatible with evolutionary principles. His suggestion that humans shape their own environment and thereby create new regularities appears even more radical than my suggestion of externalization. It certainly deserves to be taken seriously. I would not go as far as Antonio **Raffone et al.** who think that internalization is not a general bio-cognitive principle because of ever-present context and niche dependencies, but it is remarkable that the commentaries that address **SHEPARD**'s claim of an evolutionary universality of internalization find it unconvincing.

## HR6. Gestalt theory

Walter **Gerbino** rightfully points out the proximity of internalization and the Gestalt principle of Prägnanz. He seems to think that both can be tested empirically. This might be possible for Gestalt principles because a whole number of them have been suggested. It is of course possible to test which principle wins and determines the percept when two or more are brought into conflict. A Gestalt principle by itself cannot be tested. Take the principle of proximity, which states that parts that are closest to one another get grouped into one perceptual object. While this principle seems to work for the three dots on the left in Figure HR3, it does not work for the same three dots hidden amongst the other dots on the right. The principle by itself has the status of an existence statement: there exist cases where near objects group together. Like internalization, it cannot be disproved. However, while internalization is a stand-alone principle, proximity is part of a "theory" that consists of a whole number of principles. They can thus be tested with regard to their relative strengths. In this context statements can be made and tested, such as the statement that "given everything else is


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

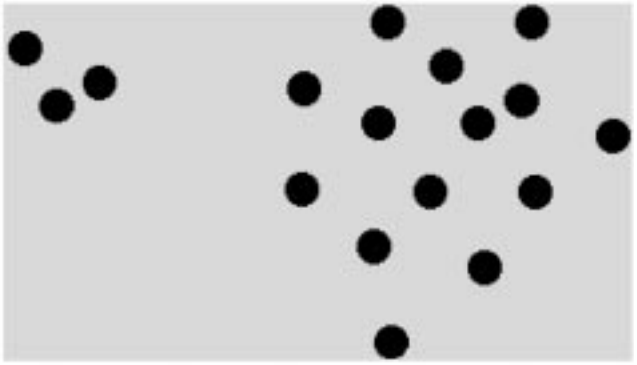


Figure HR3. Grouping by proximity does not always work

the same, proximity is on average a stronger predictor for grouping than is identity of shape." Consequently, as problematic as Gestalt theory is for a falsificationist, testable hypotheses can be derived from it. By going to just one principle, internalization, this opportunity has been forsaken.

## HR7. What regularities are internalizable?

Helen **Intraub** suggests a new candidate for an internalized regularity, namely spatial continuity. She tests this candidate in a rather ecological fashion. This could be an important piece in the mosaic of evidence testing internalization. Spatial continuity, as operationalized by boundary extension tasks, can be found in pictures and in real world scenes. Observers typically remember a larger field of view than they actually saw. The suggestion that this memory distortion reflects an internalized regularity is interesting because it is not obvious whether spatial continuity qualifies for internalization. The reason for this doubt is rooted in the usual requirement for internalization that the regularity in question acts in the world independent of the observer. Something can only be said to be internalized if it is not already a part of the beholder. This certainly holds for gravity. However, our visual boundary, as Ernst Mach (1886/1922) so aptly showed, is a condition of perception, just as is our sensitivity range to light of certain wave lengths. The statement that we have internalized light to become visible only when its wavelengths are between 300 and 800 nm, is nonsensical.

David **Jacobs**, Sverker **Runeson**, **and** Isabell **Andersson** make a similar argument when they urge us to differentiate between the constraints under which vision operates, and internalization. Hence, we can distinguish three categories of regularities that constrain what we perceive: (1) external regularities that cannot be internalized, for instance, the physical characteristics of light. Statements to the effect that we have internalized the fact that light enters the eye, or that light travels in straight lines, are nonsensical; (2) external regularities that can be internalized (Shepardian regularities); and (3) perceptual regularities that fall out of our sensory makeup (e.g., the recency effect in memory, priming, the fact that light of 400 nm typically appears as blue). It will not be helpful to call categories 1 and 3 "internalized." That is, a Shepardian regularity (category 2) has to be internalizable and it has to contribute to solving the underspecification problem. As interesting as **Intraub**'s evidence for boundary extension is in its own right, I am afraid it falls into category 3 and does not qualify as a Shepardian regularity.

The issue whether perception necessarily merely relies on environmental regularities, as **Jacobs et al.** like to think or whether some regularities have been internalized relates back to my initial distinction between instantiation and reprogrammable representation. Jacobs et al. obviously dislike the latter and find ecological psychology to be inconsistent with the notion of representation.

## HR8. The developmental perspective

Horst **Krist** and Bruce **Hood** separately recommend that we consider regularities that are being acquired during infancy and childhood. In those cases where such regularities emerge in the process of maturation, this seems reasonable. However, both Krist and Hood call for broadening the concept of internalization to include all sorts of representations acquired during childhood. Although such an empiricist interpretation of internalization is certainly possible, it makes the concept indistinguishable from learning and even less accessible to empirical testing. In my opinion, we should do everything but "loosen the criterion," as **Hood** suggests, because the concept is already hard to come by. The looser the criterion the less useful the theory. The nice developmental example of a straight-down belief that he shows supports internalization. But what do we do with a concept that has both supporting and refuting evidence? What we need is a tighter criterion. Or we admit that we are dealing with a non-statement that may be useful to organize our thinking but that ultimately evades empirical testing. In the latter case we should stop designing experiments that have the goal to test the concept of internalization.

**Krist**'s suggestion to define internalized regularities in a task-dependent manner seems directly in contradiction to his suggestion of broadening the definition. Task-dependence takes some abstraction away from the concept, and narrows it down to a specific realm of application. This appears very similar to my attempts at operationalizing the concept, albeit with the additional need to find independent predictors that specify to which tasks it should and should not apply. If we cannot find those predictors and instead have to enumerate the tasks for which internalization holds, the concept loses its appeal dramatically.

**Krist** is concerned that my differential treatment of perception and prediction has led me to interpret everything as evidence against internalization. This is a peculiar argument and I think it is founded in a difference of opinion on what constitutes the terrain for internalized constraints. I regard the application of internalization to well-specified visual events as nonsensical. We do not need internalization to explain the percept of a straight path when in fact a computer animation of straight path was presented to the observer. In general, correct performance in a rich sensory environment is not a test case for internalization. Internalization lends itself to explain judgments that are without sufficient stimulus support. In the paper-and-pencil version of the C-shaped tube problem, where no such support is provided, the error of predicting curved paths is compatible with kinematic geometry. However, there is also an indirect way to assess internalization by the use of relative comparisons of sufficiently specified events: the comparison of "relative naturalness" judgments of different, but



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

equally well-specified, events. If such events are judged as natural or anomalous, this judgment has to be based on some representation of how things should behave. Internalized constraints are candidates here. Thus, the incorrect predictions made on the C-shaped tube task speak for internalization of curved paths, while the correct judgments of visually animated versions of the task indicate that knowledge of the correct straight paths has been "internalized." Evidence for and against internalization of kinematic geometry can be found using the same naïve physics task. I believe that the contradiction between predicted outcomes and perceptual judgments is worrisome and that it is incompatible with the notion of universal internalized regularities. Would it help to make internalization task-dependent and claim that curved paths have only been internalized when the nature of the task is cognitive rather than perceptual? I do not think so, unless we want to generally limit internalization to cognitive tasks.

**Krist** is a radical internalizationist when he equates the external regularity with a perceptual regularity. Indeed, because we have eyes, the retinal image size relates to the size and distance of an external object. This relationship, however, is not a Shepardian regularity. It certainly is not external. That's why I only mentioned it in passing. It is likewise meaningless to talk about our ability to make time-to-contact judgments of approaching objects (*tau*, e.g., when catching a ball) as if it were internalized. Regularities can be internalized, faculties cannot. This relates to the issue of the interaction between the perceiver and her environment addressed below.

Does a developmental approach to internalization make sense? We could operationalize and say that the earlier an infant witnesses knowledge about a given regularity, the more deeply it must be internalized. The nice evidence by Spelke et al. (1992) supports such an endeavor. Based on their findings we would conclude that continuity and object solidity are more deeply internalized than gravity and inertia. But by saying this, have we really said anything other than "knowledge about continuity and solidity develops earlier than knowledge about inertia and gravity"? We always end up with the same problem: unless strictly operationalized, internalization remains a non-statement. We do not add anything to what we already know by using the concept. Also, the more fundamental the regularity, the less appropriate the notion of internalization. The fact that infants stop in front of a visual cliff can prompt us to state that depth has been internalized. But again, depth is not external to the observer. In other words, Gibson's (1979) affordances, such as potential injury, solidity, and so on, appear different from what constitutes an internalization. First, there is no ambiguity to the situation. The visual system has no need to resort to a generalized default. Second, neither depth nor solidity are universal. There are shallow and gaseous objects. Some creatures can walk through rain, others cannot. Failing to recognize a solid object as such does not necessarily mean that the concept of solidity has not been internalized. The infant may have "chosen" the wrong internalized regularity, as I have demonstrated for the case of horizontality.

Let us come back to **Krist**'s suggestion about task-dependent constraints. Adults and infants expect things to fall down in some cases (see also **Hood**'s commentary). We therefore have internalized the regularity of falling or gravity in those tasks where it applies. In a similar vein, we can find some evidence for the internalization of every single regularity that comes to mind. However, such a conceptualization is not very useful unless we could claim that gravity has been internalized across the board. I have provided counterexamples. The existence of examples and counter examples for every regularity that has been proposed should set off an alarm. It certainly sets off my falsificationist alarm. Demanding a task-dependent use of internalization and focusing on supporting examples does not do a service to **SHEPARD**'s theory, rather, it immunizes it against criticism. It is perfectly fine to immunize a concept, as long as we admit to it. Once immunized, we no longer need to waste our time on empirical tests. We now have to judge whether the concept is useful in guiding our thinking or whether it is time to think of more fruitful concepts.

## HR9. Does internalization operate in normal perception?

**SHEPARD** claims that internalized constraints, as observable in imagery or in cases of stimulus paucity, represent universals that are also at work in normal vision. In fact, this is why we went through all the trouble of investigating apparent motion trajectories. Now, if we generously interpret the entire data on the topic as inconclusive, what does this entail for the underlying argument that the internalized constraints also work in normal vision? I contend that even conclusive evidence for internalization under stimulus paucity – if there were such evidence – would not allow strong conclusions about normal vision because of its fundamental difference. In normal vision we recall (or directly perceive) some surplus meaning that is important to us when the object is unambiguously perceived and the stimulus is sufficient for its identification. For instance, the knowledge that cars are made of steel becomes effective when the car is in front of our eyes. On the other hand, the fact that objects move in circular paths becomes effective when the path is *not* in front of our eyes. It could well be the case that the system is only resorting to internalized knowledge when at a loss, similar to inferences we draw about what must have happened when we see a crumpled object. Observers are remarkably good at inferring the causal history of stationary objects based on their shape (Leyton 1989). **SHEPARD**'s hypothesis that unconstrained perception or imagery reveals the workings of normal perception is certainly intriguing but it is not supported by empirical evidence.

## HR10. A pragmatist view of internalization

Peter **Todd** and Gerd **Gigerenzer** implicitly question the theoretical content of internalization by calling it a metaphor. The choice of the term "metaphor" is questionable but the underlying pragmatism deserves consideration. They go on to compare internalization to other "metaphors," in particular to Simon's concept of bounded rationality, which posits a loose relationship between environment and perceiver, namely that of "scissors," where one blade stands for the task environment and the other for the computational abilities of the decision-maker (Simon 1990). **Todd & Gigerenzer** opt for the pragmatic choice based on pure convenience, just like picking up a convenient tool from a toolbox. Since every metaphor conveys a little truth we


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

should use what works best (mirror for perception, scissors for cognition). This pragmatic approach has the advantage that it is indifferent to the theoretical content of any given position and thus need not worry about finding one theory that explains the data satisfactorily. However, the position appears almost fatalistic as to the goal of science. It seems to imply that an approximation to the true state of affairs is futile. While a pragmatist position has many virtues when it comes to personal belief, it may not be very conducive to empirical science. This is as different from my admittedly unfashionable falsificationist position as can be. It can only be surpassed in pessimism by Feyerabend's (1975) anarchistic theory of science, which claims that any theory goes, regardless of evidence or practicability.

## HR11. Externalization

By suggesting externalization I only intended to sketch an alternate theory to show that it might be just as plausible as the concept of internalization even if it runs into the same problems. With this move I wanted to criticize the directionality of **SHEPARD**'s hypothesis. Like the now outdated intromission ideas about vision, the impression of external laws into our visual system appears lopsided. We need to consider the other side, our need for motor action, and ask how it impinges on our percepts. Rather than thinking in terms of intake, it might be beneficial to merely look for commonalities between the environment and our perception. Our body is just as much an environment as are the objects outside. In the evolution of a perceptual or cognitive system, both world and self are players. We may even need the additional mental player suggested by Rainer **Mausfeld**. For now, however, I gladly share the physicalistic trap. Adding aspects of the perceiver to the physicalist account may be sufficient to explain regularities that are not internalizable, such as being edible.

In reply to Margaret **Wilson**, my argument that a hypotheses must apply to a potentially indefinite set of phenomena holds for internalization more so than for Gestalt laws. The latter form an ensemble and therefore do not come with the same claim of universality. While it is just an annoyance that each individual Gestalt law is as hard to falsify as it is persuasive, internalization lacks the benefit of being part of a larger whole. Thus, watered-down versions of internalization, such as statements that some regularities have been internalized and influence perception only some of the time, are not able to benefit from similar synergy. We simply end up with a non-statement. Such a non-statement is persuasive exactly because it could not possibly be wrong. Its truth is inversely proportional to its explanatory value. Exactly the same holds – of course – for the principle of externalization. I indeed conjured up externalization as a thought-provoking challenge, not as a full fledged counter theory, as Mary **Kaiser** aptly points out. It is only true in the watered-down version that some physical body constraints are reflected in our perception some of the time. In this context, the concept of an emulator that is mapping body movements onto perceptual objects, as proposed by **Wilson**, is very intriguing. It might indeed offer a venue to elaborate on the externalization idea and to turn it into something more substantial than a challenge to internalization.

## HR12. Dissolving the concept of internalization

Douglas **Vickers** and Martin **Kurthen** both suggest dissolving internalism. Their radical suggestions deserve to be explored further and could eventually lead to an alternative concept that is much superior to **SHEPARD**'s notion of internalization. **Vickers'** suggestion of a "concrete instantiation of . . . most general principles" is certainly compatible with the demand for falsifiability. General principles are easy to falsify and therefore highly desirable. As far as I understand his generative transformation, it does not need to make claims about internalization but is content with identifying general rules of perception, which might even include the body constraints I had in mind. It might indeed be better to formulate visual abilities in terms of rules for prediction and thus circumnavigate the issue of internalization vs. externalization. Thereby, the concept of internalization is dissolved or overcome.

Martin **Kurthen** on the other hand, wants to overcome internalization by inverting the explanatory process. Rather than using internalization (as representation) to explain cognition, he suggests that cognition has to explain representation. Although this radical approach sounds very intriguing, it might not work in cases where the explanans is a general principle and the explanandum a very specific behavior. For instance, when internalization of curves is used to explain a very specific apparent motion percept, a general principle is used to explain a particular fact. It is hard to imagine how the particular fact could be used, in turn, to explain the general principle. If we say the fact embodies the principle, the direction of explanation from the general to the specific is maintained.

A different attempt to dissolve the concept of internalization might be even more radical although it has been in co-existence with cognitive theories of perception for a long time. It is the direct realism put forth by **Wilson & Bingham**. They deny that **SHEPARD**'s model is about human perception because dynamics can be directly perceived and representation is not needed altogether.

## HR13. Conclusion

In summary, the beauty and the frustration of the concept of internalization lie in its lack of content. To postulate internalized constraints as a structural form of mental representations amounts to a non-statement, albeit an elegant and fascinating one. Without further specification, the idea is appealing because it is true by definition. And much of the argument is about definitions. The lack of precision has the great advantage that the internalizationist does not necessarily have to make up her/his mind and side with or against representationalism, ecological realism, constructivism, and so on. This advantage also frustrates the critic. The critic does not get anywhere if she argues against internalization from a particular theoretical stance. Evidence against particular representations does not put a dent in the concept of internalization, and evidence against the general concept is impossible. The commentaries have shown that the concept of internalization continues to fascinate many, while others think it might have outlived its usefulness. This lack of agreement indicates that a satisfactory unifying theory does not exist. Therefore, we have to be pluralistic and content ourselves with narrower theories that are – desirably – falsifiable.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

## Kubovy & Epstein's Response

## Internalization: A metaphor we can live without

Michael Kubovy and William Epstein

*Department of Psychology, University of Virginia, P.O. Box 400400, Charlottesville, VA 22904-4400.* **kubovy@virginia.edu**
**we7v@virginia.edu** **http://www.virginia.edu/~mklab/**

**Abstract:** We reply to Gerbino, Heschl, Hoffman, Jacobs et al., Kurthen, O'Brien & Opie, Todorovič, and Wilson. Several issues are clarified. We concede little.

**Gerbino** wishes that we had included the Gestalt psychologists in the landscape we mapped out in our target article, Section 1.2 ("Locating Shepard"), in which we situated **SHEPARD** with respect to Helmholtz, Transactionalism, Rock, Marr, and Gibson. Had we had more room to be expansive we would have done just what he recommends, but probably not as ably as he has done.

Nonetheless, we find ourselves wondering about **Gerbino**'s assertion that "Shepard is very close to Gestalt theory." He makes three points in support of this assertion. *First*, that "exploring percepts as end-products of the equilibrium between external and internal forces has been *the* Gestalt strategy." Perhaps a history of our field will reveal that this is a unique contribution of the Gestalt psychologists, but even if it does, currently the idea could hardly be more widespread in cognitive psychology. So we are not sure why Gerbino thinks that this is a clear differentiating feature of the Gestalt approach. *Second*, that **SHEPARD** and Gestalt theory both distinguish "between formal (e.g., minimal coding) and processing simplicity." But as we see it, for SHEPARD the disposition of the visual system to prefer simple solutions is a reflection of internalized geometric constraints, whereas for the Gestalt psychologists it is an inherent property of dynamical systems and not a reflection of the environment. *Third,* Gerbino wonders why we take issue with the wide-spread use of the innocuous metaphor of internalization. The answer to this question is the heart of our article, which boils down to two points: (1) in some cases, some metaphors can be misleading, and are therefore not innocuous, and (2) SHEPARD's use of the term internalization may mislead. But we do agree with Gerbino that SHEPARD probably disagrees on this matter with the Gestalt psychologists. Shepard's notion that motion perception is governed by regularities internalized over the phylogenetic history of the organism could be considered antithetical to the Gestalt psychologists' emphasis on self-organizing processes in the brain.

We're not sure why **Gerbino** and **Heschl** think that we object to the introduction of evolutionary theory into perception. This is certainly not our position in general. There is no doubt that in the course of evolution, organisms change internally, as Heschl points out. But Heschl is proposing a rather weak form of a theory of internalization, whereas we are objecting to the strong form adopted by **SHEPARD**. In our target article, Section 2.2 ("Questioning Internalization") we asked whether – in the case of motion perception – the conditions for invoking evolutionary theory have been satisfied, and we concluded that they hadn't. The situation may be different in the case of color vision. We did not address SHEPARD's work in this domain, and we should have said so explicitly in our article. (A persuasive example is Mollon 1995).

In contrast, **Kurthen** takes our critique of **SHEPARD**'s internalism to be halfhearted. Did we really say, as Kurthen claims, that "the notion of internalization results from the unconscious application of misleading metaphors in Lakoff and Johnson's (1990) sense"? We weren't proposing an analysis of SHEPARD's mind, but pointing to the importance of understanding the seductiveness of metaphors, and their danger. Because – as Lakoff and Johnson have shown – metaphors come in clusters, the use of one metaphor may lead naturally to the use of others in the cluster, as we have proposed.

There is much in **Jacobs et al.**'s commentary with which we agree. But we would not like to be misunderstood regarding the motivation for our analysis. We were motivated by general principles, and not a wish to promote the Ecological Approach to perception. We are certainly not abjuring theories that propose hidden processes in perception, if that is what Jacobs et al. mean by internal-entities theories.

**O'Brien & Opie** criticize us for arguing that the conditions we laid out in our target article, section 2.1.1("Kinematic geometry as a model"), could never be realized. But that is not what we intended: we meant to say that they have not been realized to date. The crux of our argument was this:

- **If** "the internalized constraints that embody our knowledge of the enduring regularities of the world are likely to be most successfully engaged by contexts that most fully resemble the natural conditions under which our perceptual/representational systems evolved" (Shepard 1987a, p. 266);
- **If** it is necessary to study motion perception under "the unfavorable conditions that provide no information about motion" (Shepard 1994, p. 7);
- **If** under favorable viewing conditions, "we generally perceive the transformation that an external object is actually undergoing in the external world, however simple or complex, rigid or nonrigid" (Shepard 1994, p. 7);
- **Then** we do not know under what circumstances kinematic geometry could have been internalized.

**Todorovič** criticizes our measurement theory approach to understanding internalization. Before we begin, a point regarding a matter on which we may not have been sufficiently clear: we were not proposing to formalize the relation between kinematic geometry and motion perception, but to propose what would have to be achieved for **SHEPARD**'s claims about internalization to be valid. In other words, it is precisely in order to show that the goal has not been reached that we proposed the formalization.

**Todorovič** correctly points out that motion is much richer than weight. It involves multiple ordered properties, such as speed, path length, etc. So, quite understandably, he asks how we propose to construct a measurement model of kinematic geometry. Asking the reader to keep in mind, that we're not proposing to undertake this construction, we will show how this can be done. Kinematic geometry is a geometric model of physical motion. This fact offers us the assurance that a measurement model can be found for speed, path length, etc. Thus, for each ordered property of physical motion a measurement model can be, and indeed has been, constructed. In other words, relation $k$ in Figure 2 in our target article (sect. 2.1.1, "Kinematic geometry as a model") is satisfied. So the richness of motion is not an obstacle to the application of measurement theory. According to the position we have staked out, **SHEPARD**'s claim regarding the internality of kinematic geometry requires that three more sets of homomorphisms need to be constructed to show that kinematic geometry is internal. (And then a further account is required to explain the process of internalization.)



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

In sum, we take the notion of internalization seriously, and then show what a heavy burden of proof this notion entails. Although we do not say that **SHEPARD**'s claim is incorrect, or even unprovable, (although we suspect it is both) we do argue that it is (1) difficult to prove, and (2) unproven. It turns out that **Todorovič** agrees with us: that is why he cites a wealth of evidence that physical motions and our perceptions of them are not closely related.

We had considerable trouble understanding **Hoffman**'s comments, perhaps because they were so condensed. We do not, however, apologize for what Hoffman took to be a philosophical stance. Where else would we have the opportunity to discuss metascience and philosophy of science other than in a journal such as *BBS*? We thought that **SHEPARD**'s use of the term internalization carried with it certain undesirable connotations, and we used all the means at our disposal to make that point.

On several further points we do not recognize our views in **Hoffman**'s encapsulation: (1) Hoffman thinks it is incoherent to claim that the $E \rightarrow S$ mapping "lacks an inverse even in the presence of recognition." We do not see how one can argue that recognition, i.e., a successful solution to the inverse projection problem, shows that the $E \rightarrow S$ mapping has an inverse. (2) Hoffman says that we believe that "perceptual laws and the intellectual activity of their discovery are unrelated." Just the opposite is true: it is precisely because we think that they are related that we worry about the use of the term "internalization." (3) We never said "that kinematic geometry is a superset of the perception of real motion rather than being a subset of dynamics of a rigid body." Rather, Figure 2 in our target article (sect. 2.1.1, "Kinematic geometry as a model") shows that kinematic geometry is a *model* of a subset of dynamics of a rigid body. (4) When we argue that the term "internalization" may lead to confusion, we do not "use the metaphor of emotion, which unfortunately is evolutionarily internalized," as Hoffman puts it. We mention attempts to purge the term "emotion" as applied to non-human species from the scientific vocabulary to show that we are neither terminological purists, nor behaviorists. (5) We are puzzled by Hoffman's assertion that emotion and artistic skill have been "evolutionarily internalized." That would imply that there was a time during which emotion and artistic skill were external, a claim we fail to understand. (6) Of course Hoffman is right: "cognitive psycholinguistics has little to do with the perception of motion." We wanted to clarify a point about scientific argumentation, not to claim that one can solve problems of the perception of motion by invoking findings in linguistics. (7) We discussed the issues of metaphors of mind not to be frivolous or to play on words (although we were a bit playful in a footnote) but because we wanted to explore as many aspects of the notion of "internalization" as we thought were relevant to the debate at hand.

**M. Wilson** claims that *internalization* is an apt metaphor, rather than a risky one, because "it invites contact with . . . [an] engineering solution [to the problem of designing a remote-controlled factory robot:] the use of an 'emulator,' a mechanism within the control system that mimics the behavior of the situation being acted upon, taking afferent [efferent?] copies of motor commands and producing predictions of what should happen." This indeed is a most instructive example of internalization. Such an emulator might indeed satisfy the four homomorphisms summarized in Figure 2 in our target article (sect. 2.1.1, "Kinematic geometry as a model"). If you build a robot that contains such a device, you know how it works, and you know therefore that such solutions might be good ones for biological systems. But that is all you know. When you're faced with an organism, you must *infer* internalization from the system's behavior under various circumstances. Unfortunately, it is difficult to show that a system is of the "emulator" type. That is the source of our doubts.

## Schwartz's Response

## Regularities in motion: Apparent, real and internalized

Robert Schwartz

*Department of Philosophy, University of Wisconsin-Milwaukee, Milwaukee, WI 53201.* **schwartz@uwm.edu**

**Abstract:** By and large the commentators express reservations similar to my own concerning Shepard's position. Some focus on the issue of ecological validity, others are skeptical of Shepard's evolutionary thesis. The assumptions and reasons for their criticisms, however, are not always ones I fully accept. In addition, several commentators, although rejecting Shepard's particular claims, support his overall approach. They offer models for motion perception and for phenomena in other domains that they feel are amenable to treatment along Shepard's lines. I examine the pros and cons of the commentators' analyses of Shepard. I also briefly evaluate aspects of their alternative proposals.

The primary goal of my target article, "Evolutionary Internalized Regularities," was to elicit clarification from **SHEPARD** on two issues: (i) the role his kinematic principle is supposed to play in the perception of real motion, and (ii) the proper way to understand the claim and significance of his evolutionary internalization hypothesis. I referred to (i) as a problem of ecological validity in part to highlight the fact that SHEPARD sought to link his ideas to those of J. J. Gibson. Given the focus of this special issue, I will limit my responses to topics related to (i) and (ii), rather than explore in any detail the commentators' own research and theories. Here, again, I am more interested in elucidating issues and positions than in rebutting claims. Since I have not seen Shepard's replies to my paper, however, my responses to the commentators may be based on misconceptions I still have of Shepard's ideas.

**Foster** is generally sympathetic to **SHEPARD**'s project, but he wishes to offer improvements. He proposes an energy minimizing principle that he believes can better handle cases of apparent motion and can be fruitfully extended to some other areas of perception. Whether Foster's model offers a more satisfactory account of apparent motion than Shepard's remains to be determined. Either way, it should be clear that developing an adequate theory of apparent motion would be no small accomplishment, as well as of considerable interest on its own. My concerns about ecological validity were not meant to challenge Shepard's model of apparent motion, but to understand better how these ideas may or may not be applied to the perception of real motion. I have analogous concerns with Foster's model. And, to the extent the principles of apparent motion Foster proposes are not characteristic of perception in everyday environments, an evolutionary internalization story would seem to



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

be less tenable in his case. Foster appreciates this problem. He says of his own account, "it may be more difficult to maintain the notion that the rules governing these phenomena are specific adaptations to properties of the world."

**Intraub** shares my concerns that **SHEPARD** may have trouble extending his apparent motion findings to ordinary perception, but she too approves of his theoretical framework. She suggests, however, that Shepard's framework may be better applied to the representation of the spatial layout. In particular, she proposes a principle of boundary extension as an internalized universal. According to Intraub, people tend to represent themselves as having seen features of the layout that lie beyond the boundaries of what they actually have observed. Intraub thinks her principle has two advantages over Shepard's: 1. Her principle clearly reflects a regularity of the experienced external world. There is always more to be seen. 2. Her principle can be tested in ecologically rich environments.

Perhaps I was not clear in my paper, if it seemed that I have any blanket reservations about testing perceptual theories in non-ecological, laboratory settings. This is not so. In vision studies, as in other fields of research, creating "ideal" conditions, using artificial devices, isolating factors that can only be separated in the lab, and employing other non-real world setups is often the best way to explore an area. Apparent motion and real motion can both be studied in and out of the lab, in rich and impoverished environments. The issue I was intent on examining was not one of testability or testing *per se.* Rather, I wished to explore the way principles said to determine the course of apparent motion might function in the perception of real motion.

**Intraub**'s own proposal, to think of her boundary extension principle as an internalized constraint, does raise some general issues about the nature of such claims. Boundary extension, after all, is just one of the many ways we "supplement" what is "given" visually. For example, filling in the blind spot, amodal occlusion completion, assuming the constitution of the back side of objects, and apparent motion, all may also be considered cases where "projections are internally generated but constrained." One question then is whether it is appropriate to consider descriptions of all these various forms of supplementation as instances of "internalized constraints"? Or could they just as well be understood as descriptions of the *kinds of supplementation* found? More specifically, I do not know what exactly Intraub wishes to imply in calling her own projection rule an "internalized constraint."

**Intraub**'s desire to treat her boundary extension principle as a case of evolutionary internalization raises further issues about the proper interpretation and evaluation of such claims of origin and development. Granted Intraub's principle reflects a real world regularity – that there is always more to be seen. Still, what can be concluded about the evolutionary and developmental forces that may have led to the internalization of her particular boundary extension principle rather than some of the other possibilities? As Intraub notes, boundary extension has its pluses and minuses; the advantage is extended prediction, the loss is a decline in accuracy. For example, a no boundary extension rule would result in more accurate representations but in less predictive coverage. In contrast, a boundary extension rule that went further than the one Intraub finds would mean more extensive coverage. The tradeoff would be less accurate representations. Is the type of boundary extension which Intraub actually finds, "just right" or just not lethal? Are we to assume it conveys an advantage? And if it doesn't, would this affect an evolutionary internalization claim? In the end, would the explanatory content and function of her principle be any different if the constraints on boundary extension were not the result of a process of evolutionary internalization?

**Heschl** comes to **SHEPARD**'s defense by attempting to counter what he labels the "intricate epistemological problem of how to deliver a sound proof for the existence of an internalization process." I, for one, do not see my concerns as especially epistemological, nor do I ask for "absolute proof." Testability and testing, as I said, were not the focus of my remarks. Nor do the qualms I expressed about Shepard's internalization claim stem from any epistemological reservations about the status of evolution as a theory of biological change and development. Indeed, read blandly, it is hard to find fault with Heschl's claim that "every existent perceptual-cognitive mechanism . . . must have originated from an evolutionary process of adaptive internalization." Problems arise when more detailed contentful hypotheses are offered purporting to explain the existence and mechanisms of particular psychological phenomena in specifically evolutionary terms. Then more must be known about the nature of the adaptive process, about earlier and later functions, about historical environments and niches, and about the biological basis for internalization. **Heschl** indicates he is aware of these issues and difficulties when offering his own reasons for not accepting the particulars of Shepard's proposal.

**Hoffman**'s claim that **SHEPARD**'s use of "anomalous perception as a window to the inner workings of the mind is in the best tradition of William James" need not be contested. As I indicated above, I see nothing wrong in using laboratory and other non-ecological setups in psychological studies. Furthermore, when it comes to biological systems, examining unusual or diseased cases, in either normal or abnormal settings, is a tried and true method for advancing knowledge. Of course, in these circumstances one must be extra careful, since it is so much easier to go wrong when extrapolating from atypical subjects and situations to normal subjects and situations. Concerns of this sort led me to seek clarification from SHEPARD about just which aspects of his apparent motion theory are or are not supposed to transfer to the perception of real motion. Hoffman allows that my interpretation of Shepard's principle as a "turnpike" theorem is valid. Nevertheless, I am not sure how Hoffman's own analysis of the principle, supplemented with his provisoes for veridical adjustments, serves to resolve the issues of ecological validity I raised.

Claims about evolution are much in fashion in psychology, and vision theory is no exception. **Hoffman** notes, correctly, that I am uneasy with certain evolutionary psychology theses. Before responding further on this topic, it may perhaps be useful to spell out several of my misgivings in a little more depth than I did in my target article.

First, in various contexts of inquiry it is not obvious *why* it is important to distinguish internalized constraints from those that are simply internal. In terms of processes, mechanisms, and effects on perception, it would seem to make little difference whether the constraint is learned; the result of past generations' incorporating an external regularity (i.e. internalized); or due to the intersection of evolutionary factors unconnected with or only accidently connected to a worldly regularity. It is important to keep in mind that these reservations about the explanatory need for an evolutionary


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

perspective are distinct from any problems I or others may have about innateness theses or the postulation of internal representations (Atherton & Schwartz 1974; Schwartz 1969; 1995). The issue here is whether an account of origins must or should play a major role in shaping and testing proposed models of conception and perception. Rejecting an evolutionary perspective, though, is perfectly compatible with being a staunch proponent of innateness and internal representation models of cognition. Here, it is instructive to note that Noam Chomsky, surely no enemy of innate "ideas" and internal representations, is in accord with this point. He has similarly challenged the relevance of evolutionary internalization theses to both his own theory of language competence and to David Marr's computational approach to vision (Chomsky 2000).

Second, it cannot be maintained a priori that the innateness of a constraint, either one environmentally driven or one more fortuitously arrived at, precludes plasticity. A principle present at birth may be alterable by experience. The constraint may function as a default setting, remaining in force only as long as the environment is not too recalcitrant. Conversely, dependence on learning does not imply plasticity. Principles acquired by learning may resist alteration by later experience. Nor can it be presumed that innate constraints provide an advantage over acquired ones because they enable quicker processing of, or reaction to, important stimuli. Learned constraints may function equally quickly and efficiently. The need to take these considerations into account makes it that much harder to pin down and evaluate claims of adaptive advantage.

Finally, the steps in the argument from external regularities in the world to claims of genetic internalization are mined with conceptual, methodological and evidential obstacles. These days one hears a lot about the "goals" of vision and the "optimal solutions" the organism is said to come up with to solve "vision problems." But the talk is often loose, presupposes questionable teleological assumptions, and frequently rests more on intuitions than on substantial empirical evidence. Without in-depth work in physiology, genetics, evolution and more, speculation about the biological origins of perceptual and cognitive functions tend to be just that, speculation. To quote **Hoffman** quoting S. Anstis "It could be that way, but is it really?"

For these reasons, I am unconvinced that the explanatory payoffs derived from forays into evolution – at least in our present state of knowledge – are sufficient to make such efforts an urgent matter for psychological research. Nor do I find any compelling reasons to assume that psychological studies that fail to take such evolutionary tasks to heart are somehow shallow. Hence, in my paper, I indicated that I did not have the same enthusiasm for, nor saw the pressing need to adopt, the evolutionary perspective that **SHEPARD** champions and many of the commentators advocate.

**Hoffman** seems to agree with my first point on evolutionary based explanations. He allows "that models for visual processing and underlying mechanisms can be formulated and tested independent of issues of origins." But, citing Carl Hempel, Hoffman goes on to claim that by doing so "*empirical import* and *systematic import* are . . . thereby lessened." Now Hoffman is correct that, as Hempel maintained, the more links a theory has to other theories the richer it is in empirical and systematic import. Hoffman fails to note, however, that Hempel was also a firm critic of functional/teleological explanations. In various publications Hempel painstakingly analyzed the problems involved in characterizing the notion of a "function." Hempel also went on to warn of the pitfalls to be faced in attempting to explain origins in terms of functions or functions in terms of origins (Hempel 1965). All this, of course, is not in any way meant to call into question the significance of the theory of evolution. Nor is it meant to question how important it is that psychological theories be well rooted in biological fact.

**Pickering** seems to share some of the concerns I have expressed here and elsewhere (Schwartz 1969; 1995) about human plasticity and about the problems associated with assuming a sharp split between ontogenetic and phylogenetic learning. His further worry that the emphasis on evolution has tended to downplay the significance of "the special cultural scaffolding that surrounds human development" is well taken. Pickering also calls attention to some of the complexities involved in evaluating adaptive advantage. As he says, closed strategies may be developmentally favorable, especially in a given niche. The tradeoff is that they make the organism less adaptable. I think the evolutionary advantage/disadvantage bookkeeping becomes even harder to compute once it is recognized that "innate" does not entail closed or fixed, and learned does not imply open or plastic. Again, this is not to put in doubt broad claims like **Heschl**'s that there is an evolutionary story to tell about all perceptual-cognitive mechanisms. But as **Pickering** indicates, when read too blandly, such proposals can be "almost tautologous."

**Gold**'s concerns with color vision do not directly engage the issues raised in my question (i) about apparent motion. The evolutionary issues Gold raises, however, are germane to my question (ii) with regard to the significance of evolutionary internalization accounts for psychological explanation. In recent philosophical and psychological discussions of color, much is often made of the supposed adaptive advantages of color constancy. Frequently, this "fact" is prominently cited to support the author's favorite theory of what colors "really" are and why humans and other species are geared to see them in the way they do. Gold's comments raise the possibility that the main advantage color vision affords is not color constancy but chromatic contrast. Gold goes on to argue that contrast depends more on the "physical properties of ecological niches as against the global properties of the illuminent." This, **Gold** believes, raises a damaging challenge to **SHEPARD**'s universalist color internalization thesis.

Now I think it would be fascinating to know the evolutionary history of color vision in both humans and other species. And I have no idea how it will turn out. Perhaps neither color constancy nor chromatic contrast conferred serious advantages. Alternatively, they might very well have been a boon to survival, whether or not they resulted from internalizing worldly regularities. Figuring all this out would, undoubtedly, require the in-depth work in physiology, genetics, evolution, and more mentioned above. Still, I do find **Gold**'s alternative evolutionary hypothesis, albeit speculative, useful for my purposes. It helps to flesh out further my concerns about the importance of a detailed internalization account to a sound explanation of the processes and mechanisms of perception. For I am not sure what major difference it would make to research and to our understanding of the nature of color vision if Gold's origin story were correct and **SHEPARD**'s wrong.

In his commentary, **Mausfeld** appears to offer a firm



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

resolution to my just reiterated doubts. He says " . . . issues of evolutionary internalization do not bear any immediate relevance with respect to perceptual theory, because here as elsewhere in biology, a satisfactory *ahistorical* account for a functional structure does not ipso facto suffer from some kind of explanatory deficit." Mausfeld also provides a diagnosis of **SHEPARD**'s supposed misconception. Mausfeld traces Shepard's error to what he calls the *physicalist trap,* the idea that our cognitive and perceptual capacities reflect the "natural kinds" or abstract realities found in the external physical world. What's more, Mausfeld believes **KUBOVY & EPSTEIN**'s criticism of Shepard and their reluctance to accept Shepard's talk of internal principles is a result of their having the same physicalist bias.

**SHEPARD** and **KUBOVY & EPSTEIN** will, I trust, respond on their own to **Mausfeld**'s charges that they have fallen victim to physicalism's blandishments. I, for one, having pluralist, non-Realist, and non-reductionist leanings, plead not guilty. I am as bothered as Mausfeld is about various physicalist assumptions in the study of perception and cognition (Schwartz, in press). Nevertheless, I cannot go along with Mausfeld's suggestion that the reservations KUBOVY & EPSTEIN express about explicit internal representations versus *as if* internal representations need be sustained by physicalist premises. Concerns about the nature of internal representations and the proper interpretation of models that employ them do not depend on falling into the *physicalist trap.* Over the years many writers, myself included (Schwartz 1969; 1984), have raised questions about appeals to the internal representation of rules in explanations of language and other competencies. And, contra Mausfeld's citing Chomsky's (2000) authority, these criticisms and worries about "psychological reality" have not had anything to do with assuming some "odd dualism." The problems and issues were and are more mundane. To cite but a few: Should talk of knowing rules be limited to cognitive competencies as Chomsky suggests, or may the rule model be appropriately applied to such non-cognitive skills as bicycle riding? Should models of visual competence or violin playing be allied then with those of language or those of bicycling? Granted that rule model talk is appropriate in a given domain, how is its use to be understood? It seems important to distinguish activities that result from actually checking moves against an explicitly encoded recipe, from cases where the same effects are accomplished without consulting explicit rules, from cases, as in many connectionist models, where the processing is not readily decomposed into subunits having meaningful representational content. Are these cases all instances of internalized rules? Why? Why not?

**Mausfeld** approvingly cites Gestalt psychology for not falling into the *physicalist trap.* In the context of the present discussion, however, I find his mention of Gestalt "internal representations" puzzling. Gestalt accounts of apparent motion and organization, especially when overlaid on their physiological models, are usually considered to be clear cases of theories that refuse to make use of the notion of an "internal representation" in any of its more cognitive guises (Epstein 1994).

In his target article, **TODOROVIČ** raised points about **SHEPARD**'s kinematic principle that dovetailed nicely with my own. In his commentary, **Todorovič** echos **Mausfeld**'s complaints in challenging the idea that there is some simple mapping of physical principles onto perceptual phenomena. Like Mausfeld, he finds **KUBOVY & EPSTEIN**, as well as SHEPARD, guilty of this mistake. These challenges to the plausibility of physical/perceptual homomorphisms do not, of course, show that perception is not veridical. Veridicality becomes a problem only for those who adopt overly strong physicalist commitments. Nevertheless, if **Todorovič** is correct about the lack of homomorphisms, it would seem to pose problems for any attempt to apply either SHEPARD's or KUBOVY & EPSTEIN's models to the perception of real motion. Todorovič does allow that there may be other ways to correlate physical motion and motion perception that may hold up better than the mappings he dismisses. He wonders, however, "if such a description [would] amount to a restatement of the already known empirical findings of motion research." His concern here is related to one I raised in discussing **Intraub**'s boundary extension principle. There the issue was whether or how Intraub's internalization claim went beyond a description of her interesting empirical discoveries.

**Hood** wishes to stress the interplay between ontogeny and phylogeny and the importance of developmental factors. These sorts of considerations lead him to be skeptical of **SHEPARD**'s evolutionary proposal. In my target article, I distinguished the claim of internalization of a principle from the claim that a principle was internal. The former is committed to a specific account of the genesis of the principle – the evolutionary incorporation of a real world regularity. The latter is compatible with the principle being learned or acquired by some other means. For the points I wished to make in my paper, I left aside the thorny issues of internal representation and what it means for a principle to be internal or for a subject to know or operate on a rule. **KUBOVY & EPSTEIN** did focus on this issue, and I have commented on the topic above and in previous writings.

**Hood** suggests that skills and patterns of action that result from "simple learning or Jamesian habit formation" do not count as internal knowledge structures, because in his account internalized structures should be difficult to adapt. I am not sure of the rationale or justification for this analysis of "internal representation." After all, it would seem that some habits are hard to break and some knowledge structures are relatively plastic. I am also not clear just what is to be counted in and counted out on Hood's criterion. Hood, for example, cites "theory theory" models of development as fitting the internal mold. But, depending on how this model is fleshed out, it may not be a suitable one, say, for the acquisition of syntax. If so, would this mean that language competence is not to be couched in knowledge structure terms? Alternatively, Helmholtz's account of vision makes heavy use of unconscious inferences based on premises acquired by association. Does this route of acquisition render Helmholtz's model – frequently designated a paradigm case of a cognitive visual theory – unsuited for Hood's internal knowledge approach?

**Hood** proposes that "[t]he straight down bias for falling objects is a candidate for internalization." He notes, though, that unlike **SHEPARD** he is not committed to the claim that the principle was internalized as a result of evolution. The question I would like to raise to Hood is: what is at stake in calling the straight down bias an "internalized principle"? Are such principles supposed to be explicitly represented or implicitly represented? Might they simply describe tendencies in our expectations? And again, which other biases in conception and perception should be thought of as in-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ternal principles? Just those that have a "theory theory" or comparable course of development?

Although critical of **HECHT**'s arguments, **Kaiser** joins a number of commentators in distancing herself from **SHEPARD**'s claims of evolutionary internalization. She too, however, tends to assimilate "the innate" with "the absolute" and "the acquired" with "the tunable." Still, inherited or learned, Kaiser is convinced that "our visual system is not unconstrained," and she is surely right. Everyone can agree that chairs and boulders do not see, and eyes having different properties and structures than our own will see the world differently. As Kaiser says "the unconstrained hypothesis is a straw man." At the same time, I do not know how Kaiser understands the notion of a "constraint." Therefore, I am not confident I know which of all the properties, structures, limits and biases of the visual system are to be characterized in such terms.

**Kaiser** allies herself with **Mausfeld**, **Todorovič** and several others in challenging the presumption that there will be any significant match up between descriptions of the environment geared to physics and those appropriate for the study of vision. As she sees it, Gibsonian considerations should have laid such views to rest long ago. Although I may not share all of Kaiser's Gibsonian views, when further pursued, her analysis should bring her to a position in accord with mine regarding both the function and internalization of **SHEPARD**'s kinematic principle.

**Wilson & Bingham** argue that **SHEPARD**'s "model is demonstrably not about human perception." They believe motion in the actual environment is better captured by dynamic principles than it is by the kinematic ones Shepard favors. Accordingly, they feel the experiments Shepard offers in support of his position are inadequate, since these do not reflect real world tasks and conditions. Wilson & Bingham's concerns then intersect with my own worries about ecological validity. What role is Shepard's principle supposed to play in real motion perception and how plausible is his internalization thesis, if the role it does play in everyday circumstances is quite minimal? Wilson & Bingham's rejection of Shepard's claims, however, goes beyond my critique. They suggest that if Shepard's kinematic principle is not sufficient for real motion perception, it is unlikely to be sufficient in impoverished situations either. I, on the other hand, did not attempt to dispute the claim that the kinematic principle might play a role in apparent motion perception. I even speculated that Shepard's principle could be operative in perceiving real motion. I suggested that it might function as a probabilistic "soft" constraint, a constraint that had either a weak influence or at least had to be overcome in real motion perception. The problem I raised for this version of Shepard's thesis was that as far as I knew the sort of empirical studies needed to sustain a soft constraint interpretation had not been carried out.

Much of **Jacobs**, **Runeson**, **and Andersson**'s comments reflect their interest and participation in early and ongoing debates over whether perception is direct or indirect. They come down on the Gibsonian direct side, and they see this as reason to dispute **SHEPARD**'s thesis. I, myself, believe much of this controversy is terminological rather than substantive, and I have attempted to explain why in Schwartz (1994a; 1996a). So, for example, **Jacobs et al.**, following Gibson, are willing to allow that constraints may be acquired – historically, at least, a step toward the indirect, Helmholtzian camp. Were they also willing to substitute the terms "premise" or "hypothesis" for that of a "constraint," I think the distance between their approach and that of certain indirect theorists might turn out to be less than typically assumed.

In several places I have noted my sympathy with those who find it difficult to understand and empirically justify assorted appeals to internally represented principles. In addition, the most cursory review of the cognition/perception literature would show many more interpretations, uses, and contradictory understandings of the idea than the two **KUBOVY & EPSTEIN** entertain (see below). In my own paper, I tried to avoid the morass, since my problems with **SHEPARD**'s thesis were pretty much independent of how this issue is settled. Likewise, I believe Shepard's proposal can be evaluated without resolving the direct/indirect controversy, which is often run together with the internal representation debate. Thus, one need not be a Gibsonian to find that the ecological validity issues **Jacobs et al.** raise mesh with those I wished to highlight in my questions (i) and (ii).

**Kurthen** probes deeply into the background, content, and widespread use of representational talk in cognition and perception. He uses his exploration to argue against **SHEPARD**'s appeal to internal principles. Some of Kurthen's analysis I agree with, some I do not. Sorting these matters out, though, would lead to issues considerably beyond the scope of this volume. Nor am I confident about how strong an internal representation claim Shepard makes, or even needs to make, in order to sustain his case. In my target article, for example, the particular claim I focused on was Shepard's contention that we tend to perceive motion in accordance with his kinematic principle. Tendency claims of this sort need not commit Shepard to any rich version of internal representation, certainly not to one requiring that the principle be explicitly encoded in symbols. The problems I explored did not depend on taking a robust representationalist reading of "internalization." I was puzzled about how Shepard wished his hypothesis regarding the existence and manifestation of his kinematic tendency to be understood, especially in environments where there was information available to "force" veridical perception.

**Kurthen** notes the distinction I draw between an evolutionary internalization claim and the weaker claim that some internal principle is at play. Internalization implies internalism but not vice versa. Like various other commentators, Kurthen is willing simply to jettison the internalization thesis. In Kurthen's case this is not surprising, since he rejects the internalism claim that his internalization thesis entails. I, on the other hand, am reluctant to dismiss or sidestep the internalization thesis, because I see it to be at the heart of **SHEPARD**'s project.

**O'Brien & Opie** attempt to rebut **KUBOVY & EPSTEIN**'s criticism of **SHEPARD**, by pointing out that the options Kubovy & Epstein consider, rule following and "as if" rule following, are not the only ones possible. There are many other ways to interpret claims about rules and rule following that lie somewhere between explicit representation and "as if" instantiation. O'Brien & Opie offer a third interpretation, functional resemblance. They believe this approach to internal representation is better positioned to fend off Kubovy & Epstein's criticisms. Given the plethora of different, often incompatible construals of the notion of an "internal representation" that are available, it is only to be expected that Kubovy & Epstein's attempt to criticize Shepard's version would leave many other options standing.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

O'Brien & Opie, though, do seem to admit that the kinematic principle, even understood in their way as a functional resemblance, probably holds only for apparent motion. Thus, my worries about ecological validity and the plausibility of Shepard's internalization thesis remain in place. This is not surprising in that I believe the problems arise as well on the weaker "tendency" reading of internal representation I examined.

It is worth noting that in a good deal of the debate over the nature of an internal representation, the idea appears to be accepted that were there actually an explicit representation of the rule or principle, an explanation of the psychological phenomena would be at hand. Unfortunately, things are not as clear cut as one might hope. In his influential book, *The Concept of Mind*, Gilbert Ryle (1949) not only pointed out the importance of distinguishing "knowing how" from "knowing that," he forcefully argued that "knowing that" did not entail or insure "knowing how." Neither being able to articulate the relevant principles of physics or the generative rules of French grammar is, by itself, sufficient to enable you to keep a bicycle upright or carry on a French conversation. You need, in addition, the requisite skills and capacities to instantiate these principles and rules. Once this competence or knowhow is on hand, though, it is no longer obvious what need or role remains for any explicit representation. Furthermore, if the rules can drop out in this way and not be missed, why assume an explicit representation was ever needed? Note, finding this Rylean analysis and line of attack persuasive does not in any way depend on falling into the *physicalist trap* or on accepting an "odd dualism."

**Krist** wishes to defend **SHEPARD**'s theses from **HECHT**'s criticism. In doing so, however, he diminishes their scope and depth. Krist is more than willing to abandon a strong evolutionary internalization claim, one I take to be central to Shepard's thesis. Thus, I assume, Krist, too, does not believe that models of vision and cognition that do not take into account origins are damagingly shallow. In addition, Krist admits that Shepard's kinematic principle does not hold for the perception of real motion. He suggests, though, that the perception of real motion may, nonetheless, be "contaminated" by such constraints. With this I agree. As I mentioned in my response to **Wilson & Bingham**, I raised the possibility in my paper. I did go on to ask for empirical evidence that would serve to show that the kinematic constraint did influence the outcome or had to be overcome in the perception of real object motion. For those wishing to defend a soft constraint idea this would seem to be an area requiring further exploration.

**M. Wilson**, like **O'Brien & Opie**, wishes to rescue **SHEPARD** from **KUBOVY & EPSTEIN** by offering a third way between explicit rule representation and "as if" rule following. Her suggestion is to elaborate the notion of an internal representation in terms of an emulator model. As should be clear by now there is no shortage of alternative interpretations of "internal representation" on the market. Be that as it may, **Wilson** seems to hold that the emulator plays its primary role in minimalist environs, such as those found in laboratory experiments on apparent motion. She believes the emulator does still function in richer environments, but in those circumstances it does not determine the content of the percept. As I just repeated in my response to **Krist**, I did not deny this may be so. At the same time, I should think a Shepard type evolutionary internalization story is much harder to advance, if the principle governing the function of the emulator only comes into play in unusual situations.

**Wilson** ends by chastising **HECHT** for not considering Gestalt principles, like that of common fate. I am not clear how much support **SHEPARD**'s approach can receive from the Gestalt school of vision. First, as **Mausfeld** points out, the Gestaltists did not view their principles as straightforward reflections of worldly regularities. They understood them as structures the organism imposes on the environment. Second, in their perceptual work the Gestaltists tried to steer clear of the mentalistic/cognitivist connotations associated with current notions of "internal representation."

**D. Schwartz** is no relative, but we do have an intellectual family resemblance. We both urge that the usual cognitive science binaries of analog/digital, propositional/imagistic, modal/amodal schemes of representation are too limited to do justice to the full range of symbol systems that play a role in cognition (Schwartz 1994b). He suggests adopting Peirce's scheme instead. For reasons I have spelled out elsewhere, I think the Peircean trichotomy of symbol, icon, and index has its problems (for example, I do not believe iconic representation can be adequately characterized in terms of resemblance). In my own work on thought and imagery (Schwartz 1981; 1996b), I have found it more helpful to make use of the richer range of syntactic and semantic distinctions Goodman (1976) elaborated in his *Languages of Art*.

Given the brief outline provided, I do not think I fully understand **Schwartz**'s indexical theory of cognition. And any attempt to assess the merits and demerits of the ideas he did present would take us far afield. In any case, I do not see how the adoption of Schwartz's indexicalist approach can actually be used to support **SHEPARD**'s perceptual internalization thesis.

## Tenenbaum & Griffiths' Response

## Some specifics about generalization

Joshua B. Tenenbaum and Thomas L. Griffiths
*Department of Psychology, Stanford University, Stanford CA 94305-2130.*
**{jbt, gruffydd}@psych.stanford.edu**

**Abstract:** We address two kinds of criticisms of our Bayesian framework for generalization: those that question the correctness or the coverage of our analysis, and those that question its intrinsic value. Speaking to the first set, we clarify the origins and scope of our size principle for weighting hypotheses or features, focusing on its potential status as a cognitive universal; outline several variants of our framework to address additional phenomena of generalization raised in the commentaries; and discuss the subtleties of our claims about the relationship between similarity and generalization. Speaking to the second set, we identify the unique contributions that a rational statistical approach to generalization offers over traditional models that focus on mental representation and cognitive processes.

Our target article argued that many aspects of generalization and similarity judgment can best be understood by analyzing these behaviors as rational statistical inferences. The approach built on **SHEPARD**'s (1987b; this volume) original theory of generalization gradients around a single point in a psychological space, reformulating it in a broader Bayesian framework that naturally extends to situations of multiple


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

examples or arbitrary, nonspatially represented stimuli. We derived a general principle for rationally weighting hypotheses or features – the size principle – and showed how it explained several phenomena of generalization not addressed by Shepard's original theory, as well as a number of phenomena of similarity judgment traditionally identified with alternative set-theoretic models (Tversky 1977).

Our position provoked a broad range of reactions, from enthusiastic support to strong opposition. Regardless of the valence of the response, one common factor appears to be skepticism: those who endorse our Bayesian approach argue that we have not gone far enough, others suggest that perhaps we have gone too far, while a few see us as simply heading in the wrong direction altogether. These skeptical reactions fall into two broad classes: those that question the correctness or the scope of our analysis, and those that question, at a more fundamental level, the value that a Bayesian treatment offers over more traditional process-oriented models. The first three sections of this response are addressed to the former class, and the fourth section to the latter.

In sections T&GR1 to 3, we attempt to clarify the content of our original proposals and discuss what we take to be the most promising directions for improving or extending upon them. Section T&GR1 focuses on the size principle, our proposal for a cognitive universal in the domain of generalization and similarity. We emphasize that the principle is not an arbitrary postulate, but rather a rational learner's response to the sampling process generating the observed examples in many natural environments. We also clarify the grounds for asserting such a principle as a "cognitive universal," with the hope of satisfying those critics who thought our original article went too far in this assertion. Section T&GR2 speaks to those commentators who feel, in contrast, that our original article didn't go far enough, in failing to address important effects of context, prototypicality, or stimulus noise on generalization behavior. We illustrate, without going into details, how some of these effects can be naturally accommodated within our framework or by simple extensions. Section T&GR3 takes up the relation between similarity and generalization, focusing on whether or not the size principle provides real leverage in explaining phenomena of similarity judgment such as the greater salience of relational features.

In the final section, we deal with more general questions about the explanatory value of our Bayesian framework. Much of the power of our approach comes from operating, like **SHEPARD**, at a level of analysis that abstracts away from many details of representation and processing, similar to the computational theory of Marr (1982) or the "rational analysis" paradigm of Anderson (1990), Oaksford and Chater (1998; 1999), and others. This approach inspired a number of theoretical objections, on the grounds that the real power to explain human generalization or similarity judgment comes from precisely those details that we are missing. We certainly do not dispute the importance of representation and process – the traditional bread and butter of cognitive psychology – but we feel that any model framed at this level leaves unanswered central questions of why the mind uses these particular representations and processes. Section T&GR4 examines this debate in more detail, articulating some of the distinctive insights that may come from studying human cognition in terms of rational statistical inference.

## T&GR1. The size principle

A core proposal of our article was a principle for weighting hypotheses in generalization, or features in a similarity computation, that we called the "size principle." The size principle holds that more specific hypotheses or features (corresponding to smaller subsets of objects) will tend to receive higher weight than more general ones (corresponding to larger subsets), by a factor that increases exponentially with the number of examples observed. Much of our target article focused on the importance of this principle for explaining the phenomena of generalization and similarity, and these arguments were the main target of many commentaries. We will clarify and elaborate upon those arguments in this section, beginning with the rational basis of the principle.

### *T&GR1.1. The origin of the size principle*

At least one commentator (**Gentner**) refers to the size principle as an "assumption" of our theory. In response, we want to emphasize that the size principle itself is not an arbitrary postulate or a starting point of our Bayesian analysis, but rather one of its key consequences. In the spirit of Shepard's (1987b) original theory of generalization and other rational approaches to cognitive modeling (Anderson 1990; Marr 1982; Oaksford & Chater 1998; 1999), we derived the size principle from one reasonable model of the structure of the learner's environment – specifically, about the process generating the observed examples of some consequential class – which we called the "strong sampling" model. Under strong sampling, the examples are drawn randomly (and with replacement) from a distribution over all positive instances of the class. For simplicity, we took this distribution to be the uniform distribution, but it need not be uniform in general (see sect. T&GR2.1 below and the commentary of **Movellan & Nelson**). For Bayesian learners who assume this version of the strong sampling model, the likelihood $p(X|h)$ of observing the examples $X = \{x_1, \ldots, x_n\}$, given that $h$ is the true consequential subset, will (for any $h$ consistent with $X$) be equal to

$$p(X|h) = \frac{1}{|h|^n} \tag{1}$$

where $|h|$ is the size of the subset $h$ and $n$ is the number of examples. The size principle is just a qualitative description of this likelihood function (Equation 1). Then, via Bayes's rule,

$$p(h|X) = \frac{p(X|h)p(h)}{p(X)}, \tag{2}$$

these size-based likelihoods combine with prior probabilities $p(h)$ to determine the posterior probabilities of the hypotheses $p(h|X)$, and thereby shape the learner's generalization behavior.

We should also clarify the differences between hypothesis size or specificity and hypothesis frequency or rarity. In critiquing our analysis, **Gentner** appears to equate the specificity of a hypothesis with the rarity of its use or production. She identifies word frequency with the specificity of a hypothesis labeled by that word, and she argues against our size principle on the grounds that hypotheses of a certain type – causal relations – "are extremely highly weighted,



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

despite being ubiquitous in human reasoning." However, the specificity of a class is directly related not to its frequency of use as a label or as a hypothesis for generalization, but to the frequency of occurrence of the objects that it covers. For instance, the class of dogs is strictly more specific than the class of mammals, which implies that the frequency of occurrence of dogs is strictly lower than the frequency of occurence of mammals (because every dog is a mammal, but not vice versa). But the frequency of the basic-level term "dog" in natural language, and probably also the frequency of the class of dogs as a hypothesis for inductive generalization, are much higher than the corresponding quantities for the superordinate "mammal" (Kucera & Francis 1967). In terms of our Bayesian framework, only specificity is relevant to the strong sampling likelihood function; frequency of use, in contrast, may often be one means of assessing the prior probability of a hypothesis.

### *T&GR1.2. The necessity of the size principle*

In addition to **Gentner**, at least one other commentator argues that the size principle is not necessary to explain some of the phenomena that we have attributed to it. **Heit** focuses on our explanations for the effects of the distribution of examples on generalization gradients in a one-dimensional continuum (described in sect. 3 of our target article). Our original discussion examined the consequences of varying the range and number of examples: all other things being equal, less variability in the set of observed examples or an increase in the number of examples should lower the probability of generalizing beyond their range. **Heit** suggests that other Bayesian models which do not instantiate the size principle – in particular, his model of inductive reasoning (Heit 1998) – can predict the effects of example variability.

First, we note that Heit's (1998) model is equivalent to a version of our Bayesian analysis that employs a different assumption as to how the examples are generated, which we referred to as "weak sampling" in our target article. Several Bayesian accounts of inductive inference in philosophy of science (Horwich 1982; Howson & Urbach 1993) are also equivalent to weak sampling. Under weak sampling, the example stimuli $X$ are chosen from some distribution that is independent of the consequential class $C$. The probability of any hypothesis $h$ is therefore independent of the precise stimulus values observed, except insofar as they fall inside or outside of $h$. We can thus define $p(X|h)$ to be 1 for any hypothesis $h$ consistent with $X$, and 0 for any other hypothesis. This definition of the likelihood function for weak sampling follows standard practice in machine learning, where the model has been studied by a number of researchers (e.g., Haussler et al. 1994; Mitchell 1997), but it may have been one source of the confusion that **Dowe & Oppy** expressed over our treatment of weak sampling.

**Heit**'s claim amounts to saying that weak sampling is sufficient to explain the effects of example variability on the shape of generalization gradients. To evaluate this claim, Figure T&GR1 contrasts the predicted generalization gradients for weak sampling and strong sampling, using the example sets shown in Figures 2 and 3 of our target article. Both models predict that the range of stimuli judged as likely to have consequence $C$ will increase as the range of examples broaden, simply because they both perfectly interpolate for all stimuli within the range spanned by the examples. We agree with **Heit** that no size principle is necessary to produce this effect; any Bayesian model whose hypothesis space corresponds to all intervals in the stimulus continuum should act this way.

However, the more interesting effects of example variability and number concern the shape of generalization outside the range spanned by the examples. This region is where true generalization occurs, as opposed to mere interpolation, and it has been the focus of most empirical studies of the effects of example variability on the extent of generalization (Fried & Holyoak 1984; Rips 1989; Stewart & Chater, submitted; Tenenbaum 1999). Figure 1 shows that outside of the examples' range, the shape of generalization gradients under the weak sampling model shows no effect of the number of examples, and essentially no effect of their range. This is because, in contrast to the strong-sampling likelihood (Equation 1), the weak-sampling likelihood shows no dependence on either of these factors.[1] We thus do not agree with **Heit**'s claim that Bayesian models lacking the size principle can naturally predict the appropriate effects of example variability on generalization gradients.

The limitations of the weak sampling model emerge most clearly in cases where a more diverse sample leads to *narrower* generalizations – a routine situation in concept or word learning (Tenenbaum 2000; Tenenbaum & Xu 2000). Consider the case of someone learning a new word in a foreign language. Hearing the word used in the presence of a passing Dalmatian might well lead the learner to generalize that word to all other dogs, but hearing it used a number of times in the presence of various Dalmatians but no other kinds of dogs should cause generalization to non-Dalmatians to decrease substantially. Tenenbaum and Xu (2000) found evidence for this sort of behavior in both child and adult word learners and showed how it could be explained naturally by a Bayesian model incorporating the size principle. But the weak sampling model, lacking the size principle, cannot appropriately sharpen broad one-shot generalizations to a more specific level after seeing several similar examples.

As we wrote in our target article, weak sampling – despite its limitations in many situations of interest – almost surely plays a role in some kinds of inductive generalization. The weak sampling model constitutes a more conservative assumption about the generative process underlying a set of examples than does the strong sampling model, and may thus be applicable in domains where the sampling process is less apparent to the learner, or more variable from situation to situation. These might include some of the inductive reasoning tasks studied by Heit (1998), and perhaps some category learning tasks, as discussed in section T&GR1.3.3 below.

### *T&GR1.3. The size principle as a cognitive universal*

Our proposal of the size principle as a cognitive universal was based upon the observation that the strong sampling model seems to be a good approximation to the true generative process operating in many natural learning settings. Following in the tradition of **SHEPARD**'s universal law of generalization and his other proposals for how minds reflect the world in which they have evolved (Shepard 1987b; this volume), we suggest that the size principle is the mental structure that reflects an important and pervasive aspect of world structure: the random sampling of consequential


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

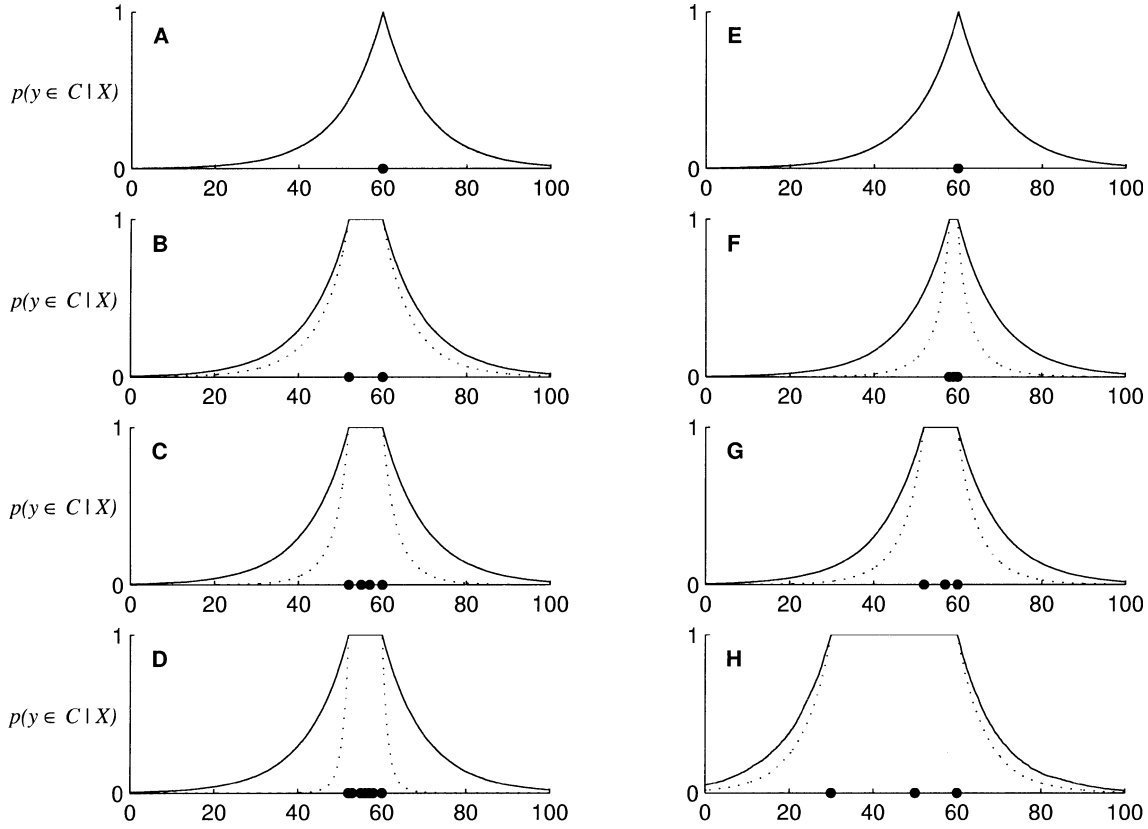



Figure T&GR1. Bayesian generalization with weak sampling. Panels A–D show that the generalization gradient is unaffected by increasing the number of examples, provided the range remains constant. Panels E-H show that the gradient is also essentially unaffected by varying the range spanned by the examples (compare the generalization gradients to the right of 60). For comparison, the dotted lines illustrate Bayesian generalization under strong sampling (equivalent to Figures 2 and 3 of our target article), showing effects of both example variability and number. To ensure that the gradients were identical for the case of a single example, the prior over the size of hypotheses was taken to be an exponential density for weak sampling and an Erlang density for strong sampling. Both priors had mean $\mu = 10$. The generalization gradients in this and all subsequent figures were computed by numerical integration, and, unless stated otherwise, used an Erlang prior over hypothesis size.

stimuli. Several commentaries take aim at this proposal, raising questions about the scope of the size principle and its potential for "universal" status. In replying to these comments, it will be helpful first to clarify exactly what we do and do not claim when we propose some cognitive principle as a universal.

**T&GR1.3.1. The scope of the size principle.** In proposing the size principle as a cognitive universal, we do claim that the principle will have broad applicability. Like **SHEPARD**'s universal law, we expect the size principle to hold across many domains of human cognition, and perhaps not only for human beings, but for any successful agent learning in an environment governed by strong sampling. This is the sense in which we believe the principle to be universal: it represents a solution to a problem that arises in many different contexts, and its effects should thus be expected to be seen in many domains. In support of the multiple-domain claim, we have found evidence that the size principle operates when people generalize in continuous multidimensional spaces (Tenenbaum 1999), when people learn simple concepts in a numerical domain (Tenenbaum 2000), when people learn concepts involving time (Griffiths & Tenenbaum 2000), and when people (both adults and four-year-old children) learn words for categories of objects (Tenenbaum & Xu 2000). There is little evidence as yet in support of the cross-species claim, although intriguingly **Baddeley et al.** suggest that, at least in some circumstances, generalization by chicks may also respect the size principle.

Despite our claims of broad applicability, we do not claim that our Bayesian framework is appropriate for every learning situation. Our analysis was based on a particular definition of the learner's task inspired by Shepard (1987b): the problem of generalizing a novel consequence from one or more examples. Many laboratory studies of category learning may not fit this definition, even if they involve some component of "generalization" broadly construed. We also do not claim that in every case where our framework applies the size principle will always hold, or that within the cases where it does hold, it will always be the dominant constraint on generalization. To reiterate a point from our target article, our theoretical analysis predicts that there should be generalization tasks for which the size principle does *not* hold, and cases in which it does hold but exerts less influence on generalization than other factors. Specifically, we should expect the size principle to be active only when the strong sampling model is valid, or more accurately, taken to be valid by the learner. Likewise, only when the prior probabilities $p(h)$ are sufficiently close to uniform should we expect the likelihood $p(X|h)$ to be the dominant factor in



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Bayes' rule – and thus the size principle to be the dominant factor in guiding generalization behavior.

Most of the commentaries criticizing our proposal of the size principle as a cognitive universal attack it on precisely these grounds, claiming that it does not hold in a particular situation, or that it may hold but be overwhelmed by the influence of other factors. We discuss each of these situations below, arguing that the putative counterexamples raised by our commentators are for the most part contained under one or more of the conditions outlined above.

**T&GR1.3.2. Variability in prior probabilities.** Several commentators, most extensively **Boroditsky & Ramscar**, describe cases where the size principle may hold but its effects on generalization seem to be outweighed by other factors. In a replication of the "all same shape" versus "all different shapes" demonstration described in our original article (Fig. 6), Boroditsky & Ramscar found that the preference for the object match over the relational match predicted by the size principle could be reversed if subjects were first given the "all same shape" array. In a structurally similar problem using letters instead of simple geometric shapes, Boroditsky & Ramscar found results consistent with the size principle, with 83% of subjects choosing the predicted letter match over the relational match. However, changing the font from Times to Chicago (which as **Boroditsky & Ramscar** describe it, makes the letters appear more similar to each other), weakened this preference so that just over half (57%) of subjects chose as predicted.

From the standpoint of our Bayesian framework, such manipulations can be thought of as changing the relative prior probabilities of the candidate hypotheses that people consider. Recall that the prior $p(h)$ expresses the learner's degree of belief that $h$ is the true subset of all stimuli with some novel consequence $C$, independent of (and logically *prior* to) observing any examples of $C$. The prior is thus the natural locus for many effects of varying feature salience independently of the particular examples observed. For instance, changing letter stimuli to a font in which letter forms are less differentiated from each other (as in the change from Times to Chicago font) will naturally decrease the prior probability of a letter-specific hypothesis, consistent with subjects' behavior. **Boroditsky & Ramscar** summarize the effect of their manipulation on the shape task in similar terms: "changing how likely it was for people to notice and represent the relational structure of the stimuli had a dramatic effect on the results of the comparison."

Far from being inconsistent with our Bayesian analysis, as **Boroditsky & Ramscar** suggest, these results provide additional support for our approach. Recall that a Bayesian learner weights hypotheses or features according to the product of prior probabilities $p(h)$ and likelihoods $p(X|h)$ (see Equation 2 above). The size principle, because it derives from a model of how the examples $X$ are sampled, is a consequence of the likelihood term, not the prior. Thus we expect that variability in both the subjects' prior probabilities and the size of hypotheses should influence behavior, just as Boroditsky & Ramscar have shown. If subjects' behavior could not be influenced by manipulating their priors, or more generally, if the size principle always explained all of the variance in any similarity or generalization experiment regardless of context or the learner's previous experience, that would represent a striking *disconfirmation* of our Bayesian analysis.

Having proposed that the size principle is a "cognitive universal," but also that it may be overruled by variability in the learner's priors, seems to leave us in a dangerously unfalsifiable position. Without some constraint on the priors, any behavior inconsistent with the size principle can be attributed to the action of priors rather than to a true failure of our proposed universal. This is plainly a chief concern of **Boroditsky & Ramscar**. Fortunately, our Bayesian framework does place one general constraint on the priors: for a particular learner in a particular context, the prior probabilites should remain constant regardless of the number or type of examples observed. In contrast, the likelihood exerts a stronger force with each successive example; recall that the size principle's preference for smaller hypotheses increases exponentially with the number of examples. This distinction provides at least one simple way to assess the relative contributions of priors and likelihoods in guiding generalization: by varying the number of examples of some novel class that subjects observe, but holding all other contextual factors constant, we expect to see the effects of the size principle emerge as the number of examples increases.

We can illustrate this analysis using a variant of **Boroditsky & Ramscar**'s third demonstration. They found that naive subjects, when asked which of 1-208-BKSDEMG or 1-911-ANALOGY was more similar to 1-615-QFRLOWY, generally chose 1-208-BKSDEMG, despite the fact that the most specific hypothesis common to 1-911-ANALOGY and 1-615-QFRLOWY, "1 in position 3, L in position 8, O in position 9, and Y in position 11" is much more restrictive than the most specific hypothesis common to 1-208-BKSDEMG and 1-615-QFRLOWY, "all different letters." Of course, this example does not really violate the size principle; **Boroditsky & Ramscar**'s explanation that the dominant features here are the distinctive features of 1-911-ANALOGY, is consistent with the fact that a hypothesis such as "contains an English word" is approximately 1,000 times more restrictive than the "1LOY" hypothesis. The interesting point for our present purposes is that, as Boroditsky & Ramscar note, people seem much less likely to consider the "1LOY" hypothesis than the far more general "all different letters" hypothesis as an appropriate generalization.

**Boroditsky & Ramscar** submit this as a counterexample to the size principle, but it is also possible that the differences in hypothesis size are just being drowned out by the extremely low prior probability that must be assigned to the "1LOY" hypothesis – presumably comparable to the prior assigned to roughly 100,000,000 other hypotheses of the same one-number-three-letters format.[2] We can distinguish these possibilities empirically by examining how similarity judgments change given multiple examples consistent with both hypotheses. If the size principle is active, it should increasingly favor the more specific hypothesis as more examples are observed. We presented 37 subjects with exactly the same task, but gave each subject either one, four, or ten examples of string stimuli that matched the "1 in position 3, L in position 8, O in position 9, and Y in position 11" hypothesis (e.g., "1-615-QFRLOWY," "1-418-MBFLODY," "1-713-NGPLOMY," and so on, presented as one string per row with corresponding string positions vertically aligned). Subjects rated the similarity of both 1-911-ANALOGY and 1-208-BKSDEMG to the set of examples, using a scale that ranged from 1 (not at all similar) to 10 (very similar). Given one example, our subjects performed like **Boroditsky & Ramscar**'s: the mean ratings were 3.3


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

and 6.5 respectively. However, as we increased the number of examples matching the "1LOY" hypothesis, 1-911-ANALOGY was judged increasingly similar to the example set while 1-208-BKSDEMG was judged increasingly dissimilar. Mean ratings respectively were 6.0 and 4.9 for four examples and 7.2 and 3.2 for ten examples, a statistically significant interaction ($F(2, 34) = 11.877$, $MSE = 6.99$, $p < 0.001$). This full reversal of the original similarity ratings argues that it was in fact prior probability differences driving subjects' initial feature weightings, and not a failure of the size principle as Boroditsky & Ramscar suggest. More generally, it illustrates how the influence of the size principle may be detected even in the presence of a strong opposing prior, by examining the dynamics of generalization or similarity as the number of examples changes.

**T&GR1.3.3. Different learning tasks, different sampling models.** While strong variability in prior probabilities typically reduces the predictive power of the size principle, we do not think that this variability accounts for all of the putative counterexamples raised by our commentators. Many of the remaining cases come from experiments on category learning, in which subjects are trained to classify stimuli into one of two or more mutually exclusive categories. **Chater et al.** describe a series of category learning experiments (Stewart & Chater, submitted) in which one prediction of the size principle – that generalization gradients will be broader for broader sets of examples – failed to hold for many subjects under certain learning conditions. **Heit** points to several previous category learning studies showing a positive correlation between category base rates and generalization rates, arguing that these results contradict a second prediction of the size principle: for a given range of examples, generalization gradients will narrow as more examples are observed. **Love**, although he does not cite any specifically contradictory studies, seems generally troubled by our lack of attention to the category learning literature.

Category learning studies differ in several important ways from the generalization tasks that were the focus of our target article, any of which could potentially account for the disparity between their results and our predictions. The most superficial differences concern the conditions of learning. Category learning experiments typically place a heavy burden on subjects' memories, presenting highly confusable stimuli one at a time over a sequence of many trials. As such, they confound the problem of generalizing beyond the examples with the problems of remembering and discriminating those examples. Because our analysis focused only on the generalization problem, it may fail to match behavior under these more uncertain learning conditions. Shepard (1987b) made a similar argument to explain the deviations from his exponential law of generalization that are often observed under similar experimental conditions. Somewhat consistent with this hypothesis, Experiment 1 of Stewart and Chater (submitted) found one predicted effect of exemplar range only when they presented all exemplars of a category simultaneously, rather than sequentially, thereby eliminating the memory burden and making discrimination of individual stimuli much easier.

The more fundamental differences between category learning tasks and our generalization tasks concern the nature of the induction problem presented to the learner. In typical category learning experiments (including those cited by **Chater et al.** and **Heit**), the categories are assumed to be mutually exclusive: during training, the learner is provided with examples of each category, and then, during testing, asked to classify new stimuli into one and only one of those categories. In contrast, our generalization tasks make no assumption of mutual exclusivity: the learner is presented with a set of examples of a particular consequential class on its own and then asked to judge the probability that different test stimuli belong to that class, independently of their membership or nonmembership in any other class.

Both of these differences – in the kinds of examples presented during training and the kinds of judgments required during testing – may make a crucial difference in how these tasks are treated in a Bayesian analysis. Depending on the nature of the examples presented during training, it may be appropriate for the learner to assume different sampling models. Depending on the nature of the judgments required during testing, computing optimal generalization probabilities may call for a somewhat different statistical model. We consider each of these differences in turn.

In our target article, we argued that the strong sampling model was most appropriate for generalizing from examples of a single class. For a learner presented with examples from two or more mutually exclusive categories, however, it is not so clear whether strong sampling or weak sampling is more appropriate. There is nothing obvious in the examples presented to tell the learner whether the examples were chosen independently of the classification to be learned (weak sampling) or sampled specifically from each of the relevant classes. In the machine learning community, well-known algorithms for classification learning have been developed based on both models (Duda et al. 2000; Mitchell 1997). In a psychological experiment, which model a subject adopts may depend on seemingly irrelevant task details, such as whether the stimuli are presented simultaneously or sequentially, or whether the alternative categories are described as two distinct classes ("X's" vs. "Y's") or positive and negative instances of a single class ("X's" vs. "not-X's").[3] Without evidence to the contrary, the more conservative weak-sampling model may thus be a rational default assumption for many human learners in many category learning tasks. This could explain the failure to find the effects of example variability and number predicted under strong sampling – but not weak sampling (Fig. 1) – in some traditional category learning tasks.[4]

Our generalization tasks require the learner to compute $p(y \in C|X)$, the probability that some new stimulus $y$ belongs to the consequential class $C$ given the observed examples $X$, but in the context of category learning, it becomes less clear how this quantity should be computed, or what it even means. Our target article (Equation 1) showed how to compute $p(y \in C|X)$ when $X$ represents a set of examples of a single class $C$, by introducing a hypothesis space of candidate subsets $h$ that could correspond to the class $C$ and averaging their predictions for the membership of $y$, weighted by their posterior probability $p(h|X)$. This prescription may not be appropriate for category learning tasks where $C$ is assumed to be one of a set of mutually exclusive categories, exactly one of which characterizes $y$. In particular, if $y$ is assumed to have been sampled from one of the mutually exclusive categories and the learner's goal in computing $p(y \in C|X)$ is to determine the probability that $y$ was sampled from $C$, then a somewhat different computational framework known as the Bayes classifier (Duda et al. 2000) is called for. In the Bayes classifier, categories are treated as



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

probability density functions and learning is modeled as estimating these densities. Several different statistical models of category learning in psychology, including parametric (Ashby 1992; Fried & Holyoak 1984), nonparametric (Nosofsky 1998), and more flexible models (Anderson 1991), can be seen as variants of the Bayes classifier. One feature of all Bayes classifiers is that they maximize the probability of responding correctly only when they bias their responses according to the relative base rates of the categories. Such a response bias would tend to increase generalization to a category as more examples of that category are observed, thereby opposing any narrowing of generalization gradients with increasing numbers of examples that might be predicted by the size principle. A rational base-rate bias may thus account for the failure to observe this particular consequence of the size principle in some traditional category learning tasks (**Heit**).

Both **Heit** and **Love** feel that our theory of generalization would have been better served by making greater contact with the extensive theoretical and empirical literatures on human category learning. However, we remain skeptical that there are many immediate connections to be made, beyond the general point that both category learning and our generalization tasks are instances of inductive inference from examples and thus can be given some kind of Bayesian treatment. Most rational analyses of category learning (e.g., Anderson 1991; Fried & Holyoak 1984; Nosofsky 1998) do not apply to our problem of generalizing from examples of a single class, and while our framework can be applied to some category learning tasks, its major predictions would change substantially and likely be difficult to distinguish from those of other theories, at least based on the existing data.

**T&GR1.3.4. Summary: What makes a cognitive universal?** Our commentators alternately criticize our framework for reaching too wide – trying to build a "universal (monolithic) theory of learning" (**Love**) that acccounts for "all of similarity and/or categorization in a simple unitary framework" (**Boroditsky & Ramscar**) – and for ignoring much important work within the broad field of categorization (**Love**, **Heit**). As we hope to have clarified in this response, neither was our intent. Our aim was to analyze a simple but universally important problem of inductive inference – generalizing from one or more examples of a novel consequence – and to show how that statistical analysis could provide rational explanations for a number of related behavioral phenomena. In particular, we did not intend – nor did we claim – to provide a rational analysis of the category learning tasks traditionally studied in many psychological laboratories, where subjects are trained to classify stimuli into a set of mutually exclusive categories and where a different statistical framework may sometimes be appropriate. We also did not claim that the main explanatory tool in our framework – the size principle – always accounts for all of the variance in generalization or similarity judgments, or that it necessarily applies in all situations where a learner must generalize from examples of a novel consequence. On the contrary, we have been quite explicit in both our target article and this response to state that under our framework, the size principle is expected to hold only under conditions of strong sampling, and to be the dominant factor in generalization or similarity judgment only when other factors instantiated in the learner's priors are relatively weak or held constant.

With all of these qualifications, how can we still maintain our proposal of the size principle as a candidate for a "cognitive universal"? **Heit** questions the broad applicability of our work to date on the grounds that "many more psychological experiments have addressed learning to distinguish one category from another" than have addressed the generalization problems we focus on. While we agree with the general goal of theoretical unification, we feel that the range of a putative cognitive universal (or any psychological theory) should be determined by the kinds of problems that organisms face in the real world, not the particular tastes of experimental psychologists. The balance of laboratory experiments notwithstanding, natural learning situations routinely involve generalization from only one or a few examples of a novel class. Learning the meaning of words in a natural language, for instance, can hardly be reduced to the problem of classifying objects into one of several mutually exclusive categories, but it has been successfully illuminated by our theoretical approach (Tenenbaum & Xu 2000).

Ultimately, the proper test of a cognitive universal cannot be its universal predictive power. First of all, there is no such thing as a universal predictor of behavior. The most we could hope for in that direction would be to establish, as clearly as possible, a set of sufficiently important tasks and sufficiently general conditions for which the predictions of a theory do apply. Shepard's (1987b) original analyis of generalization scores well in this regard, and we have tried to apply the same standards in our own work. More fundamentally, although predicting behavior is a good test of a theory's leverage, it is not our real interest as scientists. For scientific purposes, a cognitive universal is valuable because it can reflect the deep reasons why minds work the way that they do. In the domain of generalization, **SHEPARD**'s exponential law and our size principle point to the role of rational statistical inference as the basis for human inductive successes.

## T&GR2. Variations on our Bayesian framework

The account of generalization in psychological spaces that we gave in our target article made a number of simplifying assumptions, such as the restriction of hypotheses to simply connected regions, the choice of a particular analytically tractable prior probability density, the assumption of examples drawn uniformly from within the consequential range, and the lack of noise in the learner's observations. These assumptions specified one particular model of generalization that follows from our framework, but many other models using the same principles are possible. Many of our commentators discuss situations that require more complex assumptions, or suggest other models that could be generated within our framework. This section briefly considers some of these possible extensions, with the aim of showing some of the range of Bayesian generalization models that are possible even within the simple domain of a one-dimensional psychological space.

### *T&GR2.1. Prototypes and typicality*

The idea that members of a class vary in their typicality is a recurrent theme in research on concept learning and categorization. The notions of prototypes and typicality have been used to explain the finding of greater generalization to intermediate novel stimuli than to previously seen examples (Posner & Keele 1968), the apparently graded mem-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

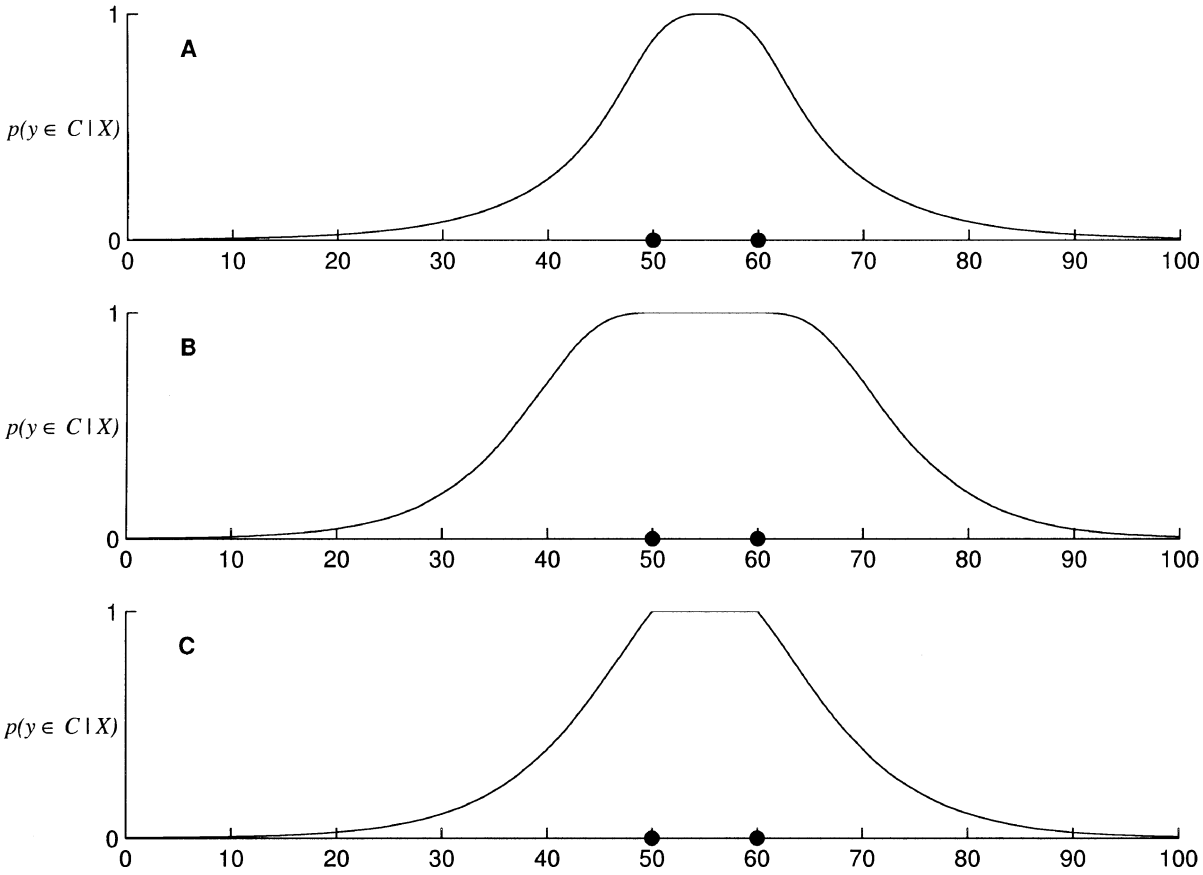



Figure T&GR2. Prototypes and typicality in generalization. Panel A shows the consequences of introducing perceptual noise, as suggested by **Baddeley et al.** The introduction of uniform noise around observations results in greater generalization to points between the exemplars than to the exemplars themselves. Panel B shows the result of using **Movellan & Nelson**'s truncated Gaussian likelihood function with two examples, leading to a highly inflected generalization gradient but no greater generalization between the examples. Panel C uses the same Gaussian form for the likelihood but integrates over all standard deviations $\sigma$, weighted according to a Gamma density, $p(\sigma) = \Gamma(2, \frac{|h|}{s})$. This extra uncertainty practically eliminates the inflection of the generalization gradient.

bership of categories (Rosch 1978), and basic asymmetries in similarity (Tversky 1977) and generalization (Feldman 1997; Rips 1975). Several of our commentators also suggest that prototypes and typicality could be and should be accommodated in our framework.

**Baddeley et al.** present data showing greater generalization to an unseen "prototype" of a class than to either of the two given exemplars (cf. Posner & Keele 1968). They suggest that this phenomenon can be captured in our framework under the assumption that the learner's observations are corrupted by some perceptual noise. In Figure 2A, we show the consequences of introducing one form of perceptual noise into our framework. Specifically, we assume that the observed value of a stimulus is not necessarily at the true value, but only within some distance of that value corresponding to the "just noticeable difference" (JND). For simplicity, we assume that the observed stimulus value is drawn from a uniform density over all points within one JND from the true value. In this figure, the points 50 and 60 are separated by 4 JNDs. The result is a clear prototype effect in the sense of Posner and Keele (1968): generalization to points between the two observed examples is higher than for the examples themselves.

**Movellan & Nelson** discuss a different way that typicality might influence generalization. They suggest that the typicality gradient of categories might be reflected in a nonuniform likelihood function $p(x|h)$, with the likelihood of observing $x$ greater for stimuli near the center of the hypothesized consequential region $h$. Using a truncated Gaussian distribution to determine $p(x|h)$, rather than the uniform distribution that our target article suggested for the strong sampling model, they show that generalization gradients may become strongly inflected in violation of Shepard's (1987b) exponential law. Interestingly, this way of incorporating typicality does not lead to the prototype effect of greater generalization to intermediate novel stimuli that was observed in the presence of stimulus noise (Fig. T&GR2B). Just as with either strong or weak sampling (Fig. T&GR1), the generalization function here always interpolates perfectly across the entire range spanned by the examples, strictly as a consequence of including all and only connected intervals in the hypothesis space. We also note that the highly inflected form of the generalization gradient under this model appears rather sensitive to the particular choice of $p(x|h)$. As the Gaussian density underlying $p(x|h)$ broadens, the degree of inflection lessens. If we do not assume knowledge of the standard deviation for the truncated Gaussian likelihood function, but instead integrate over all values of the standard deviation weighted by a suitable density function, the resulting generalization gradients have hardly any inflection at all (Fig. T&GR2C). (**SHEPARD**'s reply in this issue makes a similar observation.)



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Yet another sense of typicality arises when the psychological space represents different stimuli within a larger category, some of which might be more typical than others, and the learner's task is to generalize a novel consequence from some category members to others. To capture this sense, we can assume that stimuli are distributed in the space according to some nonuniform probability density $f(x)$, with greater values of $f(x)$ corresponding to more typical stimuli. The density $f(x)$ may reflect objective environmental frequencies, the learner's subjective beliefs, or some combination of the two. We then adopt a variant of strong sampling: instead of assuming examples drawn from a uniform distribution over the consequential class, we take them to be sampled independently and identically according to the density $f(x)$. This yields the likelihood function (for any $h$ consistent with $x$):

$$p(x|h) = \frac{f(x)}{\int_{y \in h} f(y)dy}. \quad (3)$$

Since the same term $f(x)$ appears in the numerator of $p(x|h)$ for all $h$, it cancels from the expression for $p(h|x)$. Consequently, this model becomes equivalent to the strong sampling likelihood $p(x|h) = 1/|h|$ with the size of $h$ set equal to its measure under $f$, $|h| = \int_{y \in h} f(y)dy$. **Movellan & Nelson** suggest that incorporating typicality will require us to reformulate the size principle, but under this model, we need only revise the way that the size of a hypothesis is measured.

This approach also seems to provide an explanation for some of the typicality-based asymmetries of generalization. In many contexts, people are more willing to generalize from a typical member of a category to an atypical one than vice versa. For example, upon learning that robins are susceptible to a particular disease, it seems plausible that eagles might also be susceptible. However, discovering that eagles are susceptible to a disease may result in less generalization to robins (Rips 1975). Figure T&GR3 shows the consequences of assuming that $f(x)$ is a Gaussian density with $\mu = 50$, $\sigma = 10$. Generalization from the more typical value 50 to the less typical value 60 (panel A) is higher than generalization from 60 to 50 (panel B), illustrating the asymmetry that commonly results from differences in this sort of exemplar typicality. Other potential sources of asymmetry in generalization were discussed in our target article.

In sum, different senses of typicality – and their different effects on generalization gradients – may be captured through different variations on our Bayesian framework. None of these model variations is intended to be the "correct" one, any more than one sense of typicality is the "correct" one. Rather, they represent different sets of simplifying assumptions designed to capture the most salient aspects of different learning situations.

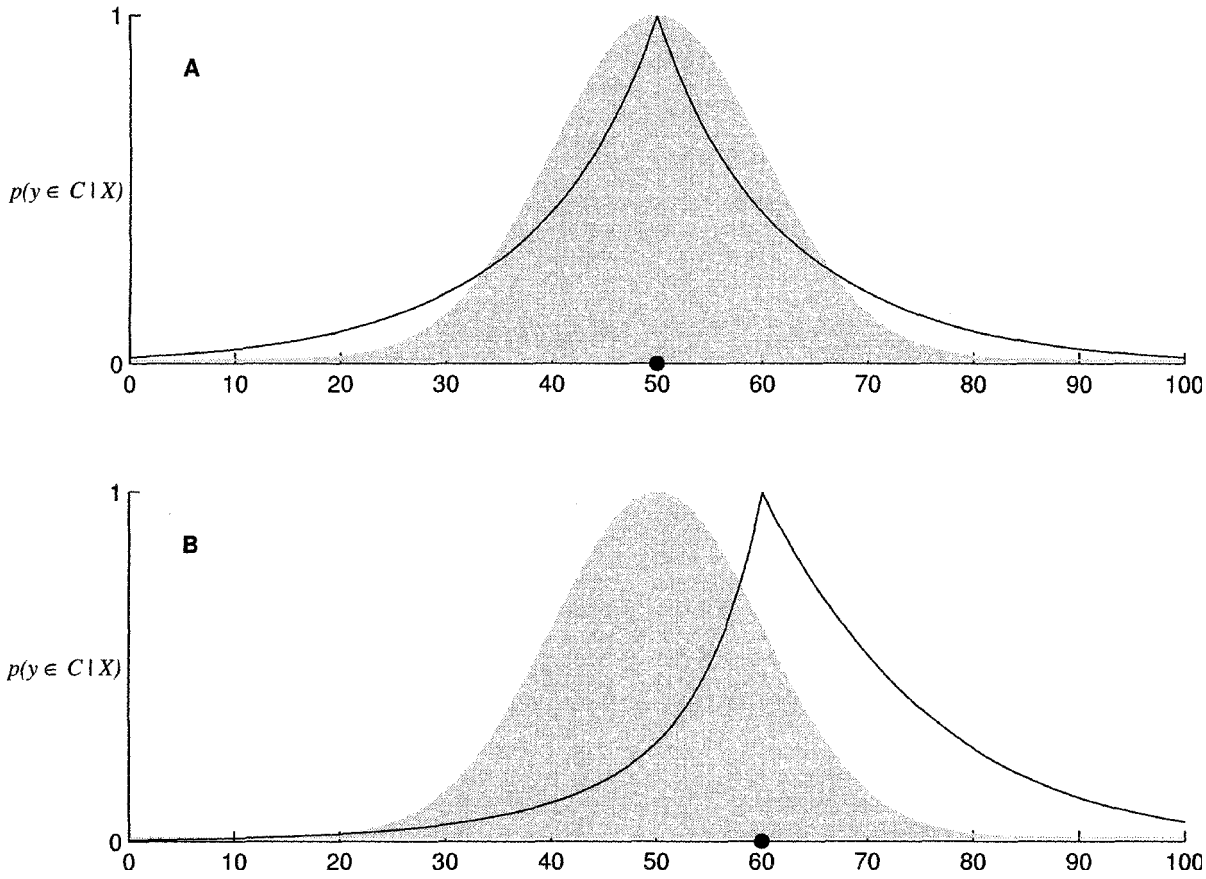



Figure T&GR3. Effect of typicality due to assuming a nonuniform sampling density for objects in a particular domain. The likelihood function within each hypothesis is proportional to the density $f(x)$, corresponding to the shaded bell-curve in the background of both panels. As a consequence, generalization is asymmetric: it is greater from a point at the mode of $f(x)$ to a point on the periphery (panel A) than vice versa (panel B).


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

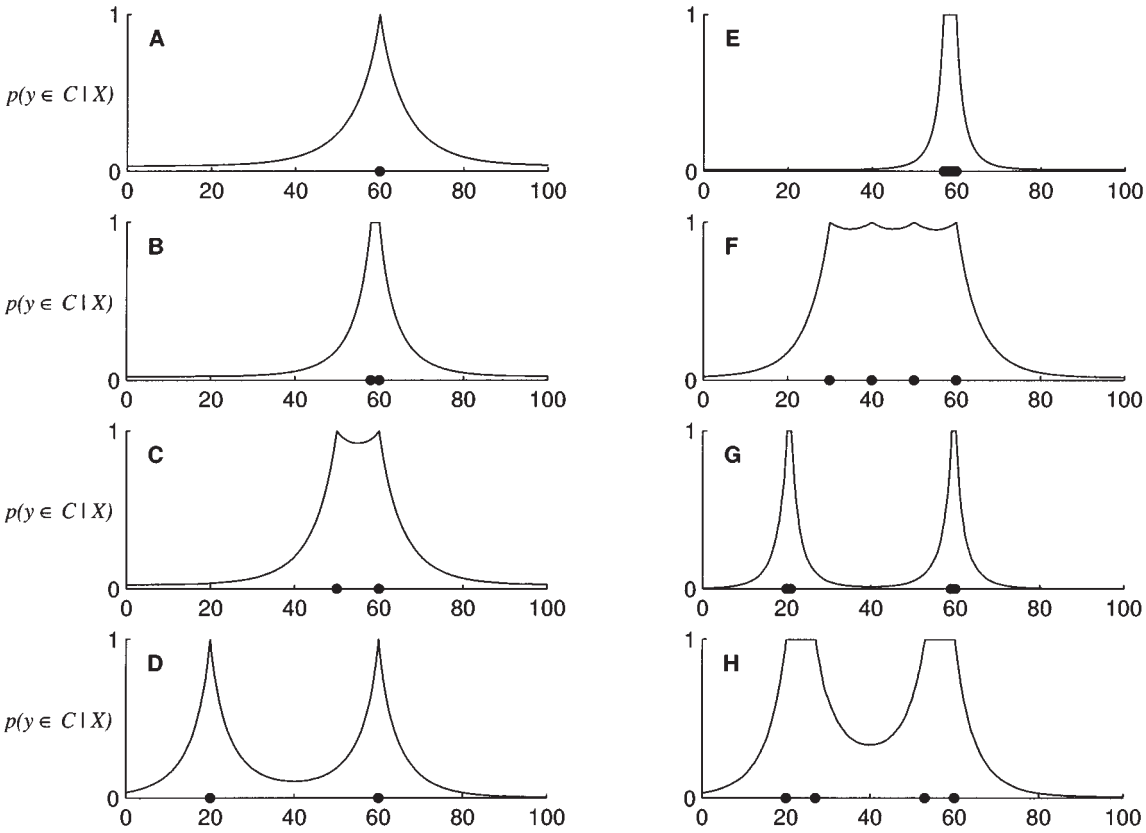



Figure T&GR4. More complex forms of generalization can be captured by more complex hypothesis spaces. Here each hypothesis consists of the union of two arbitrary intervals on the real line, which can be either connected or disconnected. To compute the prior probability of a particular hypothesis, the centers of its two component intervals are assumed to be drawn independently from a uniform density over the continuum, and the interval sizes drawn from an exponential density (with mean $\mu = 10$). This leads to more subtle effects of the distribution of examples on the shape of a learner's interpolations and extrapolations. Panel A shows generalization from a single example. The remaining panels show the consequences of different patterns of examples, as discussed in the text.

### *T&GR2.2. More complex hypothesis spaces*

The extensions described above all focus on modifications of the likelihood function $p(X|h)$. Other phenomena of generalization are more naturally captured by changes in the learner's hypothesis space $\mathcal{H}$. In particular, several phenomena of interest to our commentators may be a consequence of adding to or deleting from $\mathcal{H}$ certain systematic families of hypotheses.

Figure 4 shows some generalization gradients that arise in a one-dimensional stimulus continuum when the basic hypothesis space assumed in Shepard (1987b) and our target article – corresponding to all intervals within the stimulus continuum – is extended to include all *pairs* of intervals. For simplicity, we have assumed that the locations and sizes of the two intervals in a given hypothesis are sampled independently. Some of these hypotheses will still correspond to simply connected regions (if the two intervals intersect), but others will correspond to disconnected regions that might reflect disjunctive consequential classes, for example, the hypothesis that the good-to-eat worms are between 20 and 30 or 50 and 60 millimeters in length.

Given one example (Fig. T&GR4A), the generalization gradient is similar to the exponential functions obtained using the single-interval hypothesis space of Shepard (1987). With multiple examples, the picture becomes more complicated in a way that depends on $\mu$, the expected value (under $p(h)$) of the size of the intervals composing the hypotheses. As long as the spacing between examples is not too great relative to $\mu$, the simply connected hypotheses are most likely to correspond to the true consequential region and generalization gradients behave as before, broadening as the range of the examples increases and narrowing as the number of examples increases. But as the distances between examples becomes significantly greater than $\mu$, the disjunctive hypotheses become more likely under the size principles and there is no longer perfect interpolation within the range spanned by the exemplars. Moreover, the extent of generalization is no longer simply a monotonic function of the exemplar range. This can be seen in Figure T&GR4: increasing the range of examples at first results in broader generalization gradients (compare panels B to C, or E to F), but increasing the range still further can cause the generalization gradients to become narrower again (compare C to D, or F to G). This effect is particularly pronounced when the examples are tightly clustered (compare panels F, G, and H). Augmenting the hypothesis space to include multiple connected regions could thus explain the lack of a simple relation between exemplar variability and extent of generalization found by **Chater et al.**, with the added prediction that low generalization outside a highly variable set of examples would be correlated with a relative lack of interpolation between the examples.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

More complex hypothesis spaces require more sophisticated machinery for controlling that complexity and preventing overfitting (i.e., undergeneralization). We agree with **Lee** that our approach is compatible with Bayesian methods for model selection, and indeed to some extent the spirit of those methods is already captured in Figure 4. Equipped with a hypothesis space that allows both simply connected and disjunctive regions, the Bayesian learner effectively infers which region structure is more probable given the data. This inference is based on a balance between two forces: the size principle, which prefers disjunctive hypotheses because they are smaller, and the prior, which prefers simply connected hypotheses because they are more likely to be generated by the union of two independently chosen intervals. In this way, the learner applies a Bayesian form of Occam's razor, only entertaining complex hypotheses if the resulting improvement in the likelihood of the data outweighs the influence of the smaller prior. This automatic constraint on model complexity is a recognized advantage of Bayesian approaches to learning (MacKay 1992).

Systematic failures of generalization or interpolation in a particular area of a stimulus space, such as the lack of interpolation between yellow and blue in the chick studies described by **Baddeley et al.**, could be the result of a concentration of "missing" hypotheses in that vicinity. A learner's hypotheses might not always include, as **SHEPARD** (1987b; this volume) suggests, all possible regions of a certain shape regardless of their locations in psychological space. Perhaps, for some reason, no consequential regions are allowed which contain a certain point, or which cross a certain line, in psychological space. In any sufficiently rich collection of stimuli, there will almost certainly be some inhomogeneities of this kind, and they may be responsible for much of the meaningful perceptual or cognitive content (Feldman 1997). For color space in particular, there is independent evidence from human psychophysics for the presence of such boundaries along the yellow-gray axis (Richards & Koenderink 1995). If similar inhomogeneities are present for chicks in similar regions of color space, that could provide one explanation within our framework for the interpolation failure reported by **Baddeley et al.**

Several of our commentators (**Pothos**, **Lee**, **Boroditsky & Ramscar**) mentioned the importance of context effects in generalization. Some aspects of contextual variation could also be the result of adding or deleting hypotheses from the learner's hypothesis space. In our target article, we considered the case of a doctor trying to determine the healthy blood levels of a certain substance, knowing only that it is a hormone produced naturally by the human body and that one healthy patient has been tested and found to have a hormone level of 60 (on some suitable measuring scale). What other levels should the doctor consider healthy? Given that the substance is naturally produced by the body, having either too much or too little could be unhealthy, so it seems

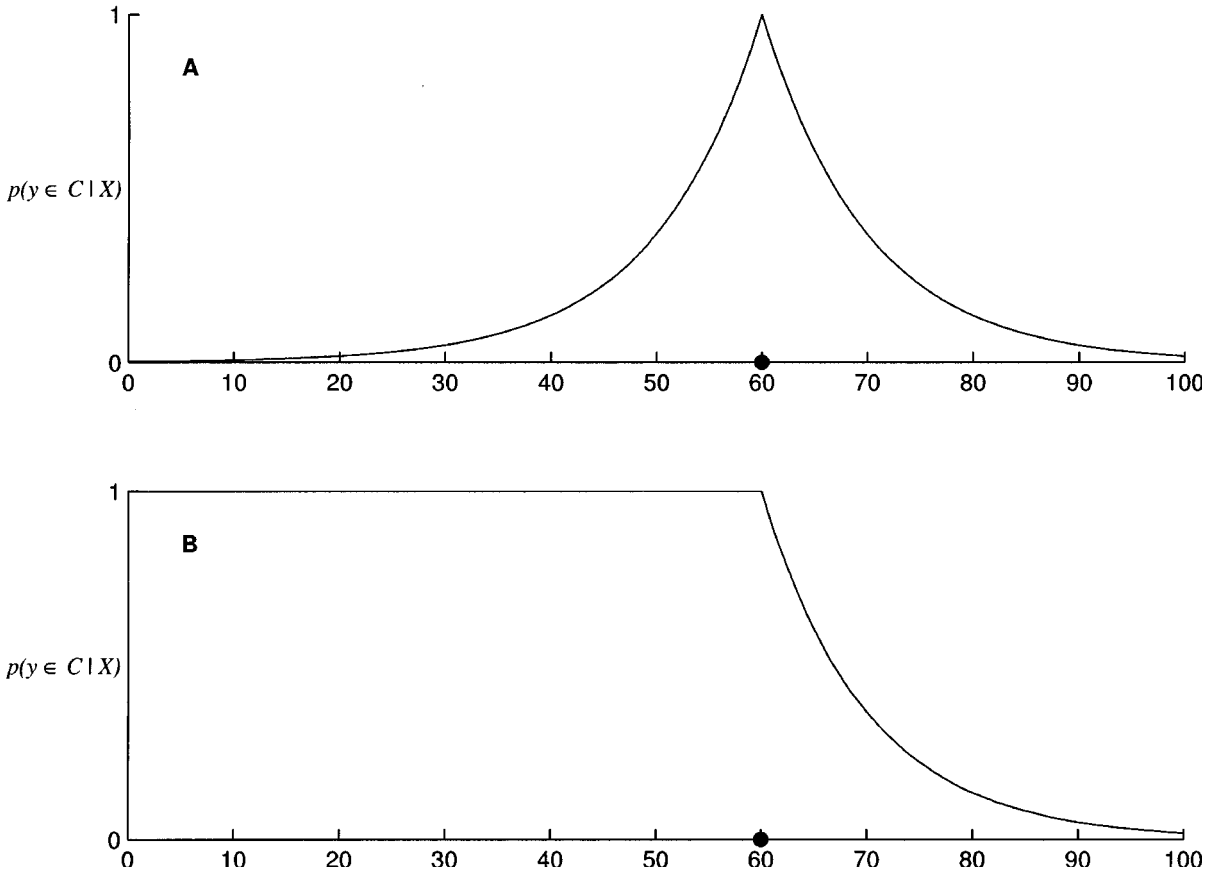



Figure T&GR5. Implementing one kind of context sensitivity in Bayesian generalization, by deleting hypotheses. Panel A shows the generalization gradient for our standard model with hypotheses corresponding to all intervals, as might be applied to determining the healthy levels of a hormone naturally produced by the human body. Panel B shows the generalization gradient with a restricted set of hypotheses, corresponding to only those intervals containing 0, as might be applied to determining the healthy levels of an environmental pollutant.


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

reasonable for generalization gradients to be approximately symmetric on either side of the one observed example (Fig. 5A). But in a different context, reasonable generalization gradients from the same stimulus value could be quite different. If, instead of being a naturally occuring hormone, the substance was believed to be an environmental pollutant, then it might be reasonable to treat as healthy any level lower than the one already observed to be healthy, but to be increasingly suspicious of higher values (Fig. 5B). This asymmetric generalization function arises naturally by deleting all $h \in \mathcal{H}$ that do not contain 0, corresponding to the reasonable belief that while different environmental pollutants may vary in their maximum healthy levels, 0 is never an unhealthy level for any pollutant. The number game described in Section 4 of our target article represents yet another context for this same stimulus set; the model there is related to the two in this section by a combination of adding and deleting hypotheses.

More generally, we can think of adding and deleting hypotheses as a special case of reweighting the priors. Deletion is equivalent to assigning zero prior probability to hypotheses that formerly received positive probability; addition is equivalent to assigning positive probability to hypotheses that used to receive zero probability. Many shifts of context may not be so severe as to actually add or delete hypotheses, but merely redistribute the prior probabilities over currently active hypotheses. Such shifts may still lead to substantially different patterns of generalization, as in our explanations for some of the manipulations of **Boroditsky & Ramscar** in section T&GR1.3.2 above.

## T&GR3. Similarity and generalization

Shepard (1987b) developed his original theory of generalization for stimuli that could be represented as points in a continuous metric space – a representation traditionally identified with the multidimensional scaling approach to modeling similarity (Shepard 1980). In extending the theory to arbitrary stimulus structures, we noted a direct equivalence to a version of Tversky's (1977) set-theoretic feature-based approach to modeling similarity. We suggested that if similarity judgments in some way reflect probabilities of generalization, then our Bayesian treatment of generalization could also offer a rational basis for some important aspects of similarity left unexplained by Tversky's (1977) original approach. Here we clarify the explanatory content and limitations of this proposal.

### *T&GR3.1. Is generalization based on similarity, or vice versa?*

Our target article argued for exploiting the relationships between similarity and generalization in two distinct ways. In Section 4.1 of the article, we conjectured that similarity judgments might depend in some way on more primitive generalization computations, and therefore that our Bayesian analysis of generalization could be useful in explaining various phenomena of similarity. But earlier in Section 4, we had proposed that people's similarity judgments could be useful in constraining Bayesian models of generalization behavior, by identifying components of the hypothesis space with the outputs of additive clustering analyses of similarity data (Shepard & Arabie 1979; Tenenbaum 1996). Some commentators (**Boroditsky & Ramscar**, **Dowe & Oppy**) found this argument to be circular, wondering how similarity judgments could depend on generalization computations if models of generalization were based on similarity judgment data.

Any apparent circularity can be resolved if we think of the relation between generalization computations and similarity judgments as causal, something like the relation between a disease and its symptoms. Because a disease tends to cause certain symptom patterns, data about those symptoms can be used to diagnose the state of the disease. Likewise, if generalization computations cause similarity judgments to behave in certain ways, then data about those similarity judgments can be used to diagnose how the underlying generalization computations work. Seen this way, the success of models of generalization built in part on analyses of similarity data (Tenenbaum 2000; Tenenbaum & Xu 2000) is not only not circular, but provides additional evidence for our claim that similarity depends causally on generalization – just as successful predictions of a disease's progression based on prior symptom observations confirms that those symptoms were in fact caused by that disease.

### *T&GR3.2. Asymmetries in similarity and generalization*

As is the case with many medical symptoms, we expect that similarity judgments may have multiple causes, with generalization computations being only one of the more important ones. For this reason, our target article did not attempt to extend our rational analysis of generalization into a precise computational model of similarity comparisons, but only to draw some insights based on "the hypothesis that similarity somehow depends on generalization, without specifying the exact nature of the dependence" (sect. 4.1). Several commentators (**Boroditsky & Ramscar**, **Gentner**) took us as actually proposing a specific model of similarity processing, which perhaps resulted in some misunderstandings about the validity of those insights.

One particular case was our suggestion that the inherently asymmetrical nature of Bayesian generalization could explain why similarity judgments are sometimes found to be asymmetric. This phenomenon has been considered central since Tversky (1977) first showed that it could be accomodated in his contrast model under suitable choices of the parameters. **Gentner** found our explanation unconvincing, on the grounds that our generalization theory's focus on the distinctive features of $x$ (in calculating the probability of generalizing from $x$ to $y$) seems to contradict the focus usually observed in asymmetric similarity judgements. However, her argument presupposes that a particular relation holds between similarity and generalization: the similarity of $y$ to $x$ must be identified specifically with the probability of generalizing from $x$ to $y$. Our target article was careful to state that the relation between similarity and generalization is complex and not reducible to any single such identification. As we wrote (sect. 4.1), "the similarity of $y$ to $x$ may involve the probability of generalizing from $x$ to $y$, or from $y$ to $x$, or some combination of those two. It may also depend on other factors altogether." Thus **Gentner**'s negative conclusion is unwarranted.

On the other hand, as **Boroditsky & Ramscar** noted, these qualifications do prevent our analysis from making definite predictions about the direction of any given asymmetric judgment. In this sense we are in a position similar to Tversky (1977), where the particular settings of free parameters in the model determine which way the asym-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

metries go, and different settings are adopted in different judgment contexts. Within our framework for studying generalization, there may also be multiple sources of asymmetries. Section T&GR2.1 above shows how asymmetries can emerge when all objects effectively have the same number of features but the features of more typical objects are weighted more highly. Even for generalization, these different asymmetries may not always point in the same direction, further complicating any attempt at predicting asymmetries of similarity.

On the subject of asymmetries, our advance over Tversky (1977) is not necessarily better predictive power, but a possible explanation of why, from a rational perspective, similarity should be asymmetric at all. Analyses of the intuitive concept of similarity often treat it as intrinsically symmetric (e.g., Bush & Mosteller 1955; Gleitman et al. 1996), which perhaps is why Tversky's (1977) original findings of asymmetries were so notable. Rather than concluding that asymmetric similarities represent some idiosyncratic quirk of how the human mind compares objects, our analysis suggests that they might – at least in part – have a rational basis in the asymmetries of generalization.

### *T&GR3.3. Relational and primitive features*

Probably the most significant advance that our analysis of similarity offers over Tversky's (1977) models is a principled explanation of some aspects of feature weighting, and how those weights might change as the number of examples in a similarity comparison change. The explanatory burden again falls on the size principle – more specific features are predicted to be, on average, more highly weighted in similarity computations. **Love** and **Gentner** take issue with our suggestion that the size principle could also explain the generally greater salience of relational features over primitive features, along with some exceptions to this tendency. They express both empirical and theoretical objections.

On the empirical front, **Gentner** offers an example of a preference for relational similarity over object similarity supposedly inconsistent with the predictions of the size principle. Given the choice between the phrases "Electrician repairing heater" or "Blacksmith having lunch" as more similar to "Blacksmith repairing horseshoe," most (15/19) of her subjects chose the relational match, "Electrician repairing heater." Gentner argues that this preference goes against the size principle since, of the two common elements, the relational term "repairing" is very frequent while the object term "blacksmith" is very rare. However, we have already cautioned against the dangers of conflating usage frequency with specificity in applying the size principle (see sect. T&GR1.1). Hypothesis specificity more properly corresponds to the number of compatible observations, which here could be assessed based on how many natural completions exist for the frames "S repairing O" or "Blacksmith V O" that are on par with the examples provided by the experimenter. Due to the multiple layers of meaning present in any natural language term, this way of measuring size will depend crucially on how the words "repairing" and "blacksmith" are used in the given examples. The word "repairing" may span a diverse set of potential meanings, but across Gentner's two examples it is used so specifically that other comparable instances of "S repairing O" would have to feature a stereotypically male tradesman repairing one of the core articles of his trade. In contrast, "blacksmith" generally sets up quite a specific context of discourse, but it is used so broadly across Gentner's two examples that comparable instances of "Blacksmith V O" could feature just about any action-object pair. The predictions of the size principle here are at best unclear, and arguably even favor the relational match that **Gentner**'s subjects chose.

To explore this further, we conducted a small study in which we manipulated the relative size of these two implicit sets. Maintaining "Blacksmith repairing horseshoe" as the phrase to which others were compared, we asked subjects to judge which of "Blacksmith pumping bellows" or "Couple repairing relationship" was more similar. We chose these stimuli to suggest a more specific interpretation of "Blacksmith V O," due to the focus on the blacksmith's characteristic metalworking activities, and a less specific interpretation of "S repairing O," through the inclusion of a nonphysical unskilled activity, with the aim of showing that specificity is in fact a driving force behind these similarity judgments. While a majority of **Gentner**'s subjects (15/19) chose the relational match, a similar majority of our subjects (13/19) chose the object match, "Blacksmith pumping bellows." This reversal, predicted under the size principle, suggests that at the least **Gentner**'s results do not contradict a role for the size principle here, and arguably even provide additional support for it.

If it does turn out that a single general factor – hypothesis size – consistently influences the relative saliencies of relational and primitive features, that would not imply that special-purpose relational processing mechanisms are unimportant. On the contrary, a subject's ability to appreciate the relative size differences for relational hypotheses (for example, in the different senses of repairing used above) seems predicated on being able to process the relations in a structurally deep way. To clarify our goals here more generally, we are not trying to dismiss the importance of process-oriented structure-mapping accounts of relational similarity (Gentner 1983; Goldstone et al. 1991), or to "subsume" them under our Bayesian framework (as **Gentner** worries). Instead, we want to understand from a rational statistical perspective some aspects of what makes this work important: why relations might be such pivotal components of mental representations, and why we might need certain kinds of relational processing mechanisms. The structure-mapping approach essentially takes these constraints as its starting point and from there develops an account of many facets of analogical processing. Our goal is to understand the rational basis of these constraints.

## T&GR4. The value of a Bayesian framework

Several of our critics questioned not only the specifics of our theory, but also whether a Bayesian analysis can yield real explanatory insights into the phenomena of interest. In this section we take up these deeper philosophical objections. The main questions concern the relative importance of traditional psychological constructs – mental representations and cognitive processes – versus the ingredients of our Bayesian framework – hypothesis spaces and rational statistical inference – in giving meaningful explanations of generalization and similarity judgment behavior.

These questions are not unique to our work, but reflect broader tensions in the field of cognitive science. The range of views in the field is well represented among our com-


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

mentators. At one extreme, **Love** feels strongly that what is needed are not rational theories based on statistical inference, but "models that account for the basic information processing steps that occur when a stimulus is encountered." At the other extreme, **Movellan & Nelson** applaud our approach and that of Shepard (1987b) as being "gloriously silent about representational and processing issues." **Love** is disturbed that our framework "makes it impossible to address important issues like whether people are interpolating among exemplars, storing abstractions, applying rules, constructing causal explanatory mechanisms, and so forth because all possibilities are present and lumped together," while **Movellan & Nelson** are relieved that "endless debates about undecidable structural issues (modularity vs. interactivity, serial vs. parallel processing, iconic vs. propositional representations, symbolic vs. connectionist models) may be put aside in favor of a rigorous understanding of the problems solved by organisms in their natural environments."

Naturally we are more inclined towards the position of **Movellan & Nelson**, but we feel that both approaches have important and complementary contributions to make in understanding the mind. The debate is not so much about whose approach is right and whose is wrong, but about different goals for a cognitive theory and what counts as a satisfying explanation of behavior. Hence, our aim in this final section is to clarify our main goals in constructing Bayesian models and to highlight the unique contributions that this approach offers beyond those of more traditional psychological accounts. We frame our statement in the form of answers to two accusations of our commentators: that we have obscured crucial issues of representation in the construction of our models' hypothesis spaces, and that we have ignored crucial processing distinctions and limitations in appealing to our theory's Bayesian inference machinery.

### *T&GR4.1. Mental representations and hypothesis spaces*

As both **Love** and **Movellan & Nelson** recognize, our framework has nothing to say directly about the nature of specific mental representations. Because our theory is computational, it does operate over representations, but those representations consist only of a description of the learner's hypothesis space with no commitment to how those hypotheses are represented in the learner's mind or brain. Several commentators (**Boroditsky & Ramscar**, **Love**, **Gentner**) feel that by working at this level of abstraction, we sidestep the real question of generalization: explaining where those hypotheses come from in the first place.

We agree that the origin of a learner's hypotheses is a fascinating and fundamental question for study. However, we disagree that it is the only question worth asking, or with the strong claim of **Boroditsky & Ramscar** that without a fully worked-out answer to that question, our framework "relies solely on hand-coded and hand-tailored representations" to "carry all of the explanatory power." Our approach follows in a long tradition of cognitive models, including Shepard (1987b), Tversky (1977), and most connectionist approaches, which take as given a representation of the stimulus domain and focuses on explaining the judgments or inferences that people make based on that representational system plus the data they observe. That our representations do not carry all of the explanatory power is demonstrated by the fact that other models of learning and inference, given these same representations, do not generally predict the phenomena of generalization and similarity that come naturally out of the rational principles of our framework.

It is worth noting two distinctive aspects of how representations are treated in our approach. First, in contrast to some other approaches such as many species of connectionism, our models make all of their representational assumptions explicit. This forthrightness may sometimes make those assumptions appear more extensive than in other approaches, but it offers the great advantage that the assumptions may be checked for psychological and scientific plausibility against all other knowledge we as modelers have available. Second, in contrast to the approaches of Shepard (1987b) and Tversky (1977) that were some of our principal inspirations, our Bayesian framework gives a rigorous account of some aspects of representational flexibility – how, for the purposes of generalization or similarity, the learner's representations of different stimulus features are reweighted in light of the observed examples.

This ability to explain some dynamics of representational change is a major advantage of Bayesian approaches (see also Heit 1998), and was at the core of most explanations in our target article. It may also provide some insight into the origins of the learner's hypothesis space. Although our target article did not try to address questions of hypothesis space construction, that does not mean that Bayesian theories are incapable of doing so (as **Gentner** suggests). Some hierarchical Bayesian models (e.g., Neal 2000) effectively allow the learner to construct a hypothesis space with complexity suited to the observed data, by dynamically reweighting different possible hypothesis spaces from an infinite "hypothesis space of hypothesis spaces," itself never explicitly represented. The basic idea is similar to the computations underlying Figure T&GR4, where a Bayesian learner effectively selects a hypothesis space consisting of one- or two-interval hypotheses based on which explanation is more probable given the data.

The main reason why we have not attempted to specify a general mechanism for constructing hypothesis spaces is that we don't believe one exists. Like **SHEPARD**, we expect that natural selection over the course of evolution may have been one force shaping hypothesis spaces for generalization, but we also see roles for processes of adaptation operating at many other time scales: development, prior learning, and even short-term priming (as discussed in sect. T&GR1.3.2). The complex end product of these processes is unlikely to be captured in models that use simple bottom-up procedures to construct representations from one static source of environmental information (e.g., Landauer & Dumais 1997; Ramscar & Yarlett 2000), placing limitations on the scope of behaviors such approaches can explain. We have tried where possible to build our representations automatically – or at least to constrain them – by applying scaling or clustering algorithms to human similarity judgments (Tenenbaum 2000; Tenenbaum & Xu 2000). Clustering algorithms are essentially unsupervised learning procedures, and as we have suggested in our target article and elsewhere (Tenenbaum & Xu 2000), they may be one important means through which human learners come to acquire some of their hypotheses for generalization. Yet we doubt that such bottom-up procedures alone will ever yield a fully satisfying account of hypothesis space construction.

In sum, we agree with **Boroditsky & Ramscar** that gen-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

eralization and similarity judgment are tremendously complex cognitive capacities, in large part because of the representational flexibility of the hypothesis spaces that allow people to bring to bear massive prior knowledge on even the most elementary stimuli. But we disagree about the possibility of finding simple principles operating within these complexities. A Bayesian model does not attempt to account for the origins of prior knowledge de novo, but merely prescribes a mapping from prior hypotheses $h$ plus observed data $X$ to rational generalization behavior. Expressing this mapping as the product of prior probabilities $p(h)$ and likelihood functions $p(X|h)$ allows us to identify simple and objective domain-general principles of learning – such as the size principle – that show up in the likelihood term, separate from the complex and subjective domain-specific representational structure that is confined to the choice of hypotheses and the prior term. Far from ignoring issues of representation and prior knowledge, or attempting to explain away their complexity, Bayesian models highlight their importance and provide us with a sharp tool for studying them. By exploring the modeling consequences of different clearly articulated assumptions about representation and prior knowledge, we can understand the functional role that these structures play in shaping generalization in a given domain.

### *T&GR4.2. Process models and rational models*

In arguing that many phenomena of generalization and similarity are best understood as instances of Bayesian statistical inference, we are – like **SHEPARD** (1987b; this volume) – clearly working in the tradition of rational approaches to cognition (Anderson 1990; Marr 1982; Oaksford & Chater 1998; 1999). This approach comes with certain theoretical commitments, as well as some noncommitments. We do not assert that any of our statistical calculations are directly implemented, either consciously or unconsciously, in human minds, but merely that they provide reasons why minds compute what they do. We are also not claiming that the human mind is any sort of general-purpose Bayesian engine, but only that certain important computations can usefully be understood in Bayesian terms. Nor in any sense are we committed to a claim that "human learning is Bayesian," any more than if we argued that certain aspects of motor control or vision could best be understood by thinking of muscles as springs, or eyes as pinhole cameras, would we be committed to the claims that "muscles are springs" or "eyes are pinhole cameras." Thus we do not find much to worry about in **Dowe & Oppy**'s concern (echoed by **Love**) that "we know from countless experiments on people that we are very far from being perfect Bayesian reasoners."

Our focus on a rational or functional level of analysis, rather than specific mechanisms of cognitive processing, is perhaps the greatest point of concern for several commentators (**Boroditsky & Ramscar**, **Love**, **Gentner**). We agree wholeheartedly that building detailed models of cognitive processes is a worthy enterprise and surely the only way to capture the full "complexity," "diversity," and "sheer variety" (Boroditsky & Ramscar) of human mental life. But for us, the primary goal of modeling is not to reproduce all – or even most – of the known phenomena of generalization or similarity. We are only interested in the phenomena to the extent that they point to the deep functional reasons why our minds work the way that they do. Hence, a finding that altering the font of a letter stimulus leads to less unanimous similarity judgments (Boroditsky & Ramscar) strikes us as interesting, but hardly a reason to abandon our search for rational constraints on similarity and start working on models of letter processing instead.

Our main reason for emphasizing rational-level explanations is that the functional questions we find most compelling are either inaccessible, or already presumed to be answered, in more mechanistic process-level models. As an illustration, **Gentner** notes that "the assumption that specific hypotheses are superior to general ones is not unique to Bayesian theories." That may well be the case. However, as we have emphasized in both our target article and this response, our analysis of generalization does not assume the size principle as a given, but explains it as a consequence of performing rational statistical inference under one reasonable model of how the examples of a class are sampled. Not only does this analysis explain why people have an important preference – an issue that other approaches simply take for granted – but it also makes additional predictions about how this preference becomes stronger as the number of examples increases, and the conditions under which the preference should hold at all. Our account of the general preference for relational features in similarity judgments was offered in the same spirit: as a rational explanation for an important cognitive preference that is typically taken as a given in standard process accounts (Gentner 1983; Goldstone et al. 1991), and which goes on to make additional predictions about how the strength of this preference will vary in different cirumstances and when it can be expected to hold.

Foremost among the "Why?" questions we want to understand is why generalization works at all. The act of generalizing from examples of a novel class is the archetypal inductive inference, which over the course of modern philosophy has variously been called a "problem" (Mill 1843), a "riddle" (Goodman 1955), a "paradox" (Hempel 1965), a "scandal" (Quine 1960) and even a "myth" (Popper 1972). How then can we possibly be so good at it? Bayesian statistics provides a principled framework for understanding human inductive successes and failures, by specifying exactly what a learner is and is not justified in concluding given certain assumptions about the environment. Process models may describe the cognitive mechanisms that mediate successful generalization, but without an analysis of the statistical logic they embody, cannot explain why those mechanisms succeed in solving a vastly underdetermined problem. Trying to explain without a statistical analysis how people can generalize so successfully from such limited evidence is, to paraphrase Marr (1982), like trying to explain how birds can fly so far on so little energy without an analysis of aerodynamics.

Recognizing the importance of statistical analysis does not diminish the importance of process-oriented models. On the contrary, Bayesian analyses often raise significant challenges for process models, in prescribing computations that are intractable if carried out exactly and therefore demand some kind of approximate algorithm. We agree with **Lee** that a fruitful line of study would be to explore "fast and frugal" heuristics (Gigerenzer & Todd 2000; Tversky & Kahneman 1974) that might, under natural learning conditions, approximate our ideal Bayesian models. Tenenbaum (1999) and Griffiths and Tenenbaum (2000) describe some preliminary steps in that direction.

In general, we would welcome more interaction between process-oriented and rational models. However, we disagree


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

with **Love**'s stance that only cognitive models incorporating "basic processing constraints (e.g., working memory limitations)" possess explanatory power. One reason to maintain a distinction between rational and process models is that, as **Boroditsky & Ramscar** point out, "there is a vast literature documenting the complexity and diversity of representations and processes involved in similarity and categorization," and who is to say which subset of these should count as "basic"? More fundamentally, even if everyone agreed on a list of basic processing constraints, we would want to know why the mind's basic processing abilities are constrained in these particular ways. The processing constraints may themselves be the phenomena we seek to understand.

Consider some of the processing constraints suggested by **Love** and other commentators: working memory limitations; the division of long-term memory into multiple systems (Cohen & Eichenbaum 1993; Squire 1992); the mechanisms by which "people combine information about stimuli" (Love & Medin 1998); and the distinction between relations and object features in similarity comparison (Gentner 1983; Goldstone et al. 1991) and memory retrieval (Gentner et al. 1993). Are all of these constraints merely accidents or compromises of evolution and brain anatomy? In what ways might they reflect (at least in part) the consequences of rational design principles? The need to control complexity in any learning system (see sect. T&GR2.2) will necessarily – and quite rationally – impose some processing limitations, as will the need to realize the system within a particular kind of physical hardware. Which of these sorts of forces is responsible for the limit of seven plus or minus two in working memory, or the division of long-term memory into separate episodic and semantic stores? Such questions may be difficult or impossible to answer definitively, but we view them as central to the mission of cognitive science as reverse-engineering the mind.

Taking for granted specific cognitive processing constraints makes these "Why?" questions unanswerable. A more productive track, analogous to the approach of Marr (1982) in vision, is to start with a computationally unconstrained rational analysis and then see what sorts of processing implications follow from it. Our target article followed this strategy in attempting to give one rational basis for the apparent distinction between relations and object features in similarity judgment, and the rational bases of working memory and long-term memory characteristics have been taken up in other recent analyses (Anderson 1990; Kareev 2000). This strategy could also suggest which processing constraints really are just accidents or "hacks," as those that consistently defy elucidation in this manner.

## T&GR5. Conclusion

Our rational analysis of generalization follows from the same basic approach that led Shepard (1987b) to his original universal law. By adopting a framework of rational statistical inference and making the minimal necessary assumptions about the structure of the learning environment, we derived a simple and broadly applicable principle for weighting hypotheses in generalization judgments – the size principle. Our framework goes substantially beyond **SHEPARD**'s original analysis, which focused on generalization from a single point in a continuous metric psychological space, to explain some of the varied forms that generalization takes in the presence of multiple examples and arbitrary representational structures. It thus promises to bring the search for universal principles of mind closer to the complexities and flexibilities of higher-level cognition.

The issue of whether simple and rational principles may characterize important aspects of human thinking and reasoning – and not just lower-level, more automatic perceptual capacities – is at the heart of the most extreme reactions from our commentators, both positive and negative. They put the prospects best in their own words. Supporters see the potential for new insights into all parts of the human mind, with **Movellan & Nelson** going so far as to declare our work "a beautiful example of the most exciting and revolutionary paradigm to hit the cognitive sciences since connectionism." Critics doubt that this style of analysis could ever yield much insight into a system as intricate and sophisticated as the human mind, with **Boroditsky & Ramscar** going so far as to warn that we are in danger of developing "a theory of spherical cows – elegant, but of little use in a world filled with cows that stubbornly insist on being cow-shaped."

All science depends on simplifications, but whether these simplifications are productive or misguided depends upon the question we are asking. If the question is how to increase the milk production of cows in Hertfordshire, then the physicist who begins his presentation to the local dairy board with "First, we assume a spherical cow . . ." is humorously off the mark. But if the question is what determines the relative speed of an animal's run or the height of its jump, then assuming spherical cows – not to mention spherical horses, spherical dogs and spherical rabbits – turns out to be quite illuminating. Just this assumption in the field of mathematical biology leads to a number of universal "size principles": scaling laws that relate body size to basic locomotory parameters for a wide range of mammalian species – not just cows in Hertfordshire (Maynard-Smith 1968). Both kinds of questions have their place in the study of cognition. Yet we believe that lasting and fundamental insights into why minds work as they do are more likely to come from thinking about them like a mathematical biologist than like a dairy farmer.

NOTES

**1.** Under weak sampling, the range spanned by the examples can have a small effect for some choices of prior, because the resulting generalization gradients correspond to the renormalization of the prior over the intervals consistent with the examples. However, this effect is negligible compared with the effect of example variability under the strong sampling model, and often goes in the opposite direction.

**2.** The prospect that some logically possible hypotheses should receive very low prior probabilities is not specific to this case, but is an essential component of rational inductive inference. As we emphasized in our target article (sect. 5), any set of observed examples will always be consistent with innumerable bizarre hypotheses – each one highly specific and therefore highly weighted under the size principle. Only the fact that there are so many of them, and the presumption that each is assigned a roughly equal share of some reasonable piece of the prior probability mass, ensures that their prior probabilities will be sufficiently low to keep them from dominating a Bayesian learner's generalizations.

**3.** Regardless of which sampling model a learner adopts for the positive examples of a class, negative examples are probably handled most naturally under weak sampling, by setting the likelihood of any hypothesis to zero if it contains one or more negative examples (Tenenbaum 1999).

**4.** In suggesting that our Bayesian framework may be compat-



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

ible, under weak sampling, with situations where the size principle does not apply, we are again mindful of **Boroditsky & Ramscar**'s concerns about falsifiability. Just as strong sampling has empirical implications, the weak sampling model also places serious constraints on behavior that can be tested empirically. Under weak sampling, observing one or more examples should not alter the relative probabilities of two hypotheses as long as they both remain consistent with the data. The corresponding empirical prediction is that observing additional examples which do not falsify any additional hypotheses should have no effect on a learner's generalization behavior.

ACKNOWLEDGMENTS

We are grateful for the support of Mitsubishi Electric Research Labs (MERL), NTT Communications Science Laboratories, and a Hackett Studentship to the second author. We thank Phillip Goff, Tania Lombrozo, and Sanjoy Mahajan for helpful comments on draft sections of this manuscript.

## Todorovič's Response

## Is kinematic geometry an internalized regularity?

Dejan Todorovič

*Department of Psychology, University of Belgrade, Belgrade, Serbia, Yugoslavia.* **dtodorovic@dekart.f.bg.ac.yu**

The author does not wish to respond to the Commentators.



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

# References

**Letters "a" and "r" appearing before authors' initials stand for target article and response, respectively**

**NOTE: All Helmholtz references/citations in the texts are listed under 'von Helmholtz' in the Consolidated References.**


Adair, R. K. (1990) *The physics of baseball.* Harper & Row. [WCH]

Agostini, T. & Galmonte, A. (1999) Spatial articulation affects lightness. *Perception and Psychophysics* 61:1345–55. [NB]

Albert, M. K. & Hoffman, D. D. (1995) Genericity in spatial vision. In: *Geometric representations of perceptual phenomena,* ed. R. D. Luce, M. D'Zmura, D. Hoffman, G. Iverson & K. Romney. Erlbaum. [aHH]

Aleksandrov, A. D., Kolmogorov, A. N. & Lavrent'ev, M. A., eds. (1969) *Mathematics: Its content, methods, and meaning.* MIT Press. [WCH]

Anderson, J. (1967) *Principles of relativity physics.* Academic Press. [DD]

Anderson, J. R. (1978) Arguments concerning representations for mental imagery. *Psychological Review* 85:249–77. [aMK]

(1990) *The adaptive character of thought.* Erlbaum. [JRM, rRNS, arJBT]

(1991) The adaptive nature of human categorization. *Psychological Review* 98(3):409–29. [EH, aRNS, rJBT]

Anderson, N. H. (1981) *Foundations of information integration theory.* Academic Press. [DWM]

(1996) *A functional theory of cognition.* Erlbaum. [HK]

Arabie, P. & Carroll, J. D. (1980) MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika* 45:211–35. [aJBT]

Arecchi, F. T. (2000) Complexity and adaptation: A strategy common to scientific modeling and perception. *Cognitive Processing* 1:22–36. [AR]

Arend, L. (1994) Surface colors, illumination, and surface geometry: Intrinsic-image models of human color perception. In: *Lightness, brightness, and transparency,* ed. A. Gilchrist. Erlbaum. [MHB]

Arend, L. & Reeves, A. (1986) Simultaneous color constancy. *Journal of the Optical Society of America A* 3:1743–51. [MHB]

Arkes, H. R., Boehm, L. E. & Xu, G. (1991) Determinants of judged validity. *Journal of Experimental Social Psychology* 27(6):576–605. [DAS]

Ashby, F. G. (1992) *Multidimensional models of perception and cognition.* Erlbaum. [rJBT]

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. & Waldron, E. (1998) A neuropsychological theory of multiple-systems in category learning. *Psychological Review* 98:442–81. [BCL]

Ashby, F. G. & Gott, R. E. (1988) Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:33–53. [EH]

Ashby, F. G. & Maddox, W. T. (1992) Complex decision rules in categorization: Contrasting novice and experienced performance. *Journal of Experimental Psychology: Human Perception and Performance* 18:50–71. [BCL]

Ashby, F. G. & Townsend, J. T. (1986) Varieties of perceptual independence. *Psychological Review* 93:154–79. [NC]

Atherton, M. & Schwartz, R. (1974) Linguistic innateness and its evidence. *Journal of Philosophy* 61:155–68. [rRS]

Atick, J. J. (1992) Could information theory provide an ecological theory of sensory processing? *Network* 3:213–51. [aHB]

Atick, J. J. & Redlich, A. N. (1990) Mathematical-model of the simple cells of the visual cortex. *Biological Cybernetics* 63:99–109. [aHB]

(1992) Convergent algorithm for sensory receptive field development. *Neural Computation* 5:45–60. [aHB]

Atmanspacher, H. & Scheingraber, H., eds. (1991) *Information dynamics.* Plenum Press. [AR]

Attneave, F. (1954) Informational aspects of visual perception. *Psychological Review* 61:183–93. [aHB]

Attneave, F. & Block, G. (1973) Apparent movement in tridimensional space. *Perception and Psychophysics* 13:301–307. [aRNS]

Avrahami, J., Kareev, Y., Bogot, Y., Caspi, R., Dunaevsky, S. & Lerner, S. (1997) Teaching by examples: Implications for the process of category acquisition. *Quarterly Journal of Experimental Psychology A* 50:585–606. [EH]

Baddeley, R. J. (1996) An efficient code in V1? *Nature (London)* 381:560–61. [aHB]

Baddeley, R. J., Abbott, L. F., Booth, M. C. A., Sengpiel, F., Freeman, T., Wakeman, E. A. & Rolls, E. T. (1997) Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proceedings of the Royal Society of London, Series B* 264:1775–83. [aHB]

Baillargeon, R., Spelke, E. S. & Wasserman, S. (1985) Object permanence in five-month-old infants. *Cognition* 20:191–208. [HK]

Bak, P. (1997) *How nature works: The science of self-organized criticality.* Oxford University Press. [PMT]

Ball, W. & Tronick, E. (1971) Infant responses to impending collision: Optical and real. *Science* 171:818–20. [HK]

Balzano, G. J. (1980) A group-theoretic description of twelvefold and microtonal pitch systems. *Computer Music Journal* 4:66–84. [aRNS]

Balzer, W., Moulines, C. U. & Sneed, J. D. (1987) *An architectonic for science: The structuralist program.* D. Reidel. [rHH]

Banach, S. (1951) *Mechanics.* Monografie Matematyczne. [WCH]

Barlow, H. B. (1959) Sensory mechanisms, the reduction of redundancy, and intelligence. In: *The mechanisation of thought processes: Proceedings of Symposiium on the Mechanisation of Thought Processes,* ed. A. M. Uttley. National Physical Laboratory, Teddington/Her Majesty's Stationery Office. [aHB]

(1961) The coding of sensory messages. Chapter XIII. In: *Current problems in animal behaviour,* ed. W. H. Thorpe & O. L. Zangwill. Cambridge University Press. [aHB]

(1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* 1:371–94. [aHB]

(1974) Inductive inference, coding, perception and language. *Perception* 3:123–34. [aHB]

(1981) Critical limiting factors in the design of the eye and visual cortex. The Ferrier lecture, 1980. *Proceedings of the Royal Society, London, B* 212:1–34. [aHB]

(1983) Intelligence, guesswork, language. *Nature* 304:207–209. [aHB]

(1989) Unsupervised learning. *Neural Computation* 1:295–311. [aHB]

(1990) A theory about the functional role and synaptic mechanism of visual after-effects. In: *Vision: Coding and efficiency,* ed. C. B. Blakemore. Cambridge University Press. [aHB]

(1992) The biological role of neocortex. In: *Information processing in the cortex,* ed. A. Aertsen & V. Braitenberg. Springer. [AR]

(1995) The neuron doctrine in perception. In: *The cognitive neurosciences,* ed. M. Gazzaniga. MIT Press. [aHB]

(1996) Banishing the homunculus. In: *Perception as Bayesian inference,* ed. D. Knill & W. Richards. Cambridge University Press. [aHB]

Barlow, H. B. & Tripathy, S. P. (1997) Correspondence noise and signal pooling as factors determining the detectability of coherent visual motion. *Journal of Neuroscience* 17:7954–66. [aHB]

Barnett, S. A. (1958) The "expression of emotions." In: *A century of Darwin,* ed. S. A. Barnett. Books for Libraries Press. [aMK]

Barrow, H. G. (1987) Learning receptive fields. In: *Proceedings of the IEEE First Annual Conference on Neural Networks,* 4, 115–21. [aHB]

Barsalou, L. W. (1985) Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11:629–54. [EMP]

(1999) Perceptual symbol systems. *Behavioral and Brain Sciences* 22(4):577–660. [DAS]

Barsalou, L. W., Huttenlocher, J. & Lamberts, K. (1998) Processing individuals in categorization. *Cognitive Psychology* 36:203–72. [EH]

Bartlett, N. R., Sticht, T. G. & Pease, V. P. (1968) Effects of wavelength and retinal locus on the reaction-time to onset and offset stimulation. *Journal of Experimental Psychology* 78(4):699–701. [DAS]

Bedford, F. L. (1999) Keeping perception accurate. *Trends in Cognitive Sciences* 3:4–11. [FLB]

(in press) Towards a general law of numerical/object identity. *Cahiers de Psychologie Cognitive/Current Psychology of Cognition.* [FLB]

Beek, P. J., Peper, C. E. & Stegeman, D. F. (1995) Dynamical models of movement coordination. *Human Movement Science* 14:573–608. [TDF]

Bell, A. J. & Sejnowski, T. J. (1995) An information maximisation approach to blind separation and blind deconvolution. *Neural Computation* 7:1129–59. [aHB]

(1997) The 'independent components' of natural scenes are edge filters. *Vision Research* 37(23):3338. [JRM]

Bernard, G. D. & Remington, C. L. (1991) Color vision in "Lycaena" butterflies: Spectral tuning of receptor arrays in relation to behavioral ecology. *Proceedings of the National Academy of Science USA* 88:2783–87. [IG]

Bertamini, M. (1996) The role of stimulus orientation in short- and long-range apparent motion. ARVO Meeting, Ft. Lauderdale, Florida, April 1996. *IOVS* 37:3. [MB]

Bertamini, M. & Proffitt, D. R. (2000) Hierarchical motion organization in random dot configurations. *Journal of Experimental Psychology: Human Perception and Performance* 26(4):1371–86. [MB]

Bertamini, M. & Smit, D. J. A. (1998) Minimization processes in apparent motion. EMPG Meeting, September 1998. *Journal of Mathematical Psychology* 42:4. [MB]

Bethell-Fox, C. & Shepard, R. N. (1988) Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance* 14:12–23. [aRNS]

Berthier, N. E., DeBlois, S., Poirier, C. R., Novak, M. A. & Clifton, R. K. (2000) Where's the ball? Two- and three-year-olds reason about unseen events. *Developmental Psychology* 36:394–401. [HK]




https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Bickhard, M. H. & Terveen, L. (1995) *Foundational issues in artificial intelligence and cognitive science.* Elsevier. [MKn]

Bingham, G. P. (1993) Perceiving the size of trees: Form as information about scale. *Journal of Experimental Psychology: Human Perception and Performance* 19:1139–61. [aHH]

(1995) Dynamics and the problem of visual event recognition. In: *Mind as motion: Dynamics, behavior and cognition,* ed. R. Port & T. van Gelder. MIT Press. [AW]

(2000) Events (like objects) are things, can have affordance properties, and can be perceived: A commentary on T. A. Stoffregen's *Affordances and events. Ecological Psychology* 12(1):29–36. [AW]

Bingham, G. P., Rosenblum, L. D. & Schmidt, R. C. (1995) Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance* 21(6):1473–93. [AW]

Bishop, C. M. (1995) *Neural networks for pattern recognition.* Clarendon Press. [JRM]

Black, T. & Schwartz, D. L. (1996) When imagined actions speak louder than words: Inferences about physical interactions. Paper presented at the Annual Meeting of the Jean Piaget Society, Philadelphia, June 1996. [HK]

Blough, D. S. (1961) The shape of some wavelength generalization gradients. *Journal of the Experimental Analysis of Behavior* 4:31–40. [arRNS]

Bogdan, R. (1988) Information and semantic cognition: An ontological account. *Mind and Language* 3:81–122. [MKn]

Boring, E. G. (1942) *Sensation and perception in the history of experimental psychology.* Irvington. [WCH]

(1951) A color solid in four-dimensions. *Psychologie Experimental* 50:154–304. [ENS]

Bottema, O. & Roth, B. (1979) *Theoretical kinematics.* Dover. [aDT]

Brancazio, P. J. (1985) Looking into Chapman's Homer: The physics of judging a fly ball. *American Journal of Physics* 53:849–55. [aHH]

Braunstein, M. L. (1994) Decoding principles, heuristics, and inference in visual perception. In: *Perceiving events and objects,* ed. G. Jansson, S. S. Bergström & W. Epstein. Erlbaum. [aHH]

Bridgeman, P. W. (1940) Science: Public or private? *Philosophy of Science* 7:36. [rRNS]

Brill, M. H. (1978) A device for performing illuminant-invariant assessment of chromatic relations. *Journal of Theoretical Biology* 78:473–78. [aRNS]

Brill, M. H. & Hemmendinger, H. (1985) Illuminant invariance of object-color ordering. *Die Farbe* 32/33:35–42. [MHB]

Brill, M. H. & West, G. (1981) Spectral conditions for color constancy via Von Kries adaptation. *Proceedings of the AIC COLOR 81 Berlin,* paper J10. [MHB]

(1986) Chromatic adaptation and color constancy: A possible dichotomy. *Color Research and Application* 11:196–204. [MHB]

Brooks, R. A. (1991a) Intelligence without representation. *Artificial Intelligence Journal* 47:139–59. [PMT]

(1991b) New approaches to robotics. *Science* 253:1227–32. [rRNS]

Brown, J. F. & Voth, A. C. (1937) The path of seen movement as a function of the vector-field. *American Journal of Psychology* 49:543–63. [rRNS]

Brown, R. O. & MacLeod, D. I. A. (1997) Color appearance depends on the variance of surround colors. *Current Biology* 7:844–49. [NB]

Bruno, N., Bernardis, P. & Schirillo, J. (1997) Lightness, equivalent backgrounds, and anchoring. *Perception and Psychophysics* 59:643–54. [NB]

Brunswik, E. (1952) *The conceptual framework of psychology.* University of Chicago Press. [JRM]

(1955) Representative design and probabilistic theory in a functional psychology. *Psychological Review* 62:193–217. [PMT]

(1956) *Perception and the representative design of psychological experiments.* University of California Press. [aHB]

Brunswik, E. & Kamiya, J. (1953) Ecological cue-validity of "proximity" and of other Gestalt factors. *American Journal of Psychology* 66:20–32. [aHB]

Buchsbaum, G. (1980) A spatial processor model for object color perception. *Journal of the Franklin Institution* 310:1–26. [aRNS]

Buchsbaum, G. & Gottschalk, A. (1984) Chromaticity coordinates of frequency-limited functions. *Journal of the Optical Society of America A* 1:885–87. [NB]

Bundesen, C., Larsen, A. & Farrell, J. E. (1983) Visual apparent movement: Transformations of size and orientation. *Perception* 12:549–58. [LMP, aRNS, aDT]

Bush, R. & Mosteller, F. (1955) *Stochastic models of learning.* Wiley. [rJBT]

Butterworth, G., Rutkowska, J. & Scaife, M., eds. (1985) *Evolution and developmental psychology.* Harvester. [JP]

Caelli, T. M., Hoffman, W. C. & Lindman, H. (1978a) Subjective Lorentz transformations and the perception of motion. *Journal of the Optical Society of America* 68:402–11. [WCH]

(1978b) Apparent motion: Self-excited oscillations induced by retarted [sic] neuronal flows. In: *Formal theories of visual perception,* ed. E. L. J. Leeuwenberg & H. Buffart. Wiley. [WCH]

Caelli, T. M., Manning, M. & Finlay, D. (1993) A general correspondence approach to apparent motion. *Perception* 22:185–92. [aHH]

Calvin, W. H. (1990) *The ascent of mind.* Bantam. [WCH]

Campbell, D. T. (1987) Evolutionary epistemology. In: *Evolutionary epistemology, rationality, and the sociology of knowledge,* ed. G. Radnitzky & W. W. Bartely. Open Court. [aHH]

Carandini, M., Barlow, H. B., O'Keefe, L. P., Poirson, A. B. & Movshon, J. A. (1997) Adaptation to contingencies in macaque primary visual cortex. *Proceedings of the Royal Society of London, Series B* 352:1149–54. [aHB]

Carlton, E. H. & Shepard, R. N. (1990a) Psychologically simple motions as geodesic paths: I. Asymmetric objects. *Journal of Mathematical Psychology* 34:127–88. [DHF, TDF, aHH, WCH, aRS, arRNS, aDT]

(1990b) Psychologically simple motions as geodesic paths: II. Symmetric objects. *Journal of Mathematical Psychology* 34(2):189–228. [aHH, WCH, aRS, arRNS, aDT]

Carroll, J. D. & Chang, J.-J. (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart-Young decomposition. *Psychometrika* 35:283–319. [arRNS]

Cassirer, E. (1944) The concept of group and the theory of perception. *Philosophy and Phenomenological Research* 5:1–35. [aMK]

Castellarin, I. (2000) Le proprietàstatistiche della riflettanza in un campione di superfici naturali. Unpublished thesis, University of Trieste, Italy.

Cataliotti, J. & Gilchrist, A. (1995) Local and global processes in surface lightness perception. *Perception and Psychophysics* 57:125–35. [NB]

Cavanagh, P. (1987) Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity and shape. *Computer Vision: Graphic Images Processing* 37:171–95. [GV]

Chang, J. & Carroll, J. (1980) Three are not enough: An INDSCAL analysis suggesting that color space has seven (±1) dimensions. *Color Research and Applications* 5:193–206. [LD]

Changeaux, J.-P. & Cannes, A. (1995) *Conversations on mind, matter, and mathematics,* trans. M. B. DeBevoise. Princeton University Press. [aMK]

Chasles, M. (1830) Note sur les propriétés génerales du système de deux corps semblables entr'eux et placés d'une manière quelconque dans l'espace; et sur le déplacement fini ou infiniment petit d'un corps solide libre [A note on the general properties of a system of two similar bodies arbitrarily positioned in space; and on the finite or infinitely small displacement of an unconstrained solid body]. *Bulletin des Sciences Mathématiques, Férussac* 14:321–26. [aRNS]

Chater, N. (1996) Reconciling simplicity and likelihood principles in perceptual organisation. *Psychological Review* 103:566–81. [aHB]

(1999) The search for simplicity: A fundamental cognitive principle? *Quarterly Journal of Experimental Psychology* 52A:273–302. [NC]

Chater, N. & Hahn, U. (1997) Representational distortion, similarity and the universal law of generalization. In: *SimCat97:Proceedings of the Interdisciplinary Workshop on Similarity and Categorization,* ed. M. Ramscar, U. Hahn, E. Cambouropolos & H. Pain. Department of Artificial Intelligence, Edinburgh University. [NC, aJBT]

Chater, N. & Oaksford, M. (1999) Ten years of the rational analysis of cognition. *Trends in Cognitive Science* 3:57–65. [rRNS, aJBT]

Chater, N. & Vitanyi, P. (submitted) Generalizing the universal law of generalization. [NC]

Chaturvedi, A. & Carroll, J. D. (1994) An alternating combinatorial optimization approach to fitting the indelus and generalized indelus models. *Journal of Classification* 11:155–70. [aJBT]

Cheng, K. (2000) Shepard's universal law supported by honeybees in spatial generalization. *Psychological Science* 11:403–408. [KC, rRNS]

Cheng, K. & Gallistel, C. R. (1984) Testing the geometric power of an animal's spatial representation. In: *Animal cognition,* ed. H. Roitblat, T. G. Bever & H. Terrace. Erlbaum. [FLB]

Cheng, K., Spetch, M. L. & Johnston, M. (1997) Spatial peak shift and generalization in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes* 23:469–81. [KC]

Chernorizov, A. M. & Sokolov, E. N. (2001) Vector coding of colors in carp bipolar cells. *Vestnik MGU* 14, *Psichologia* N 1. (in press). [ENS]

Chomsky, N. (1965) *Aspects of the theory of syntax.* MIT Press. [rRNS]

(1980) Rules and representations. *Behavioral and Brain Sciences* 3:1–61. [LTM]

(1986) *Language and problems of knowledge: The Managua Lectures.* MIT Press. [aJBT]

(2000) *New horizons in the study of language and mind.* Cambridge University Press. [RM, rRS]

Chomsky, N. & Miller, G. A. (1963) Introduction to the formal analysis of natural languages. In: *Handbook of mathematical psychology, vol. II,* ed. R. D. Luce, R. R. Bush & E. Galanter. Wiley. [aMK]

Churchland, P. S. & Sejnowski, T. J. (1992) *The computational brain.* MIT Press. [SE]

Clark, A. (1993) *Sensory qualities.* Clarendon. [LD]




https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Clark, A. & Grush, R. (1999) Towards a cognitive robotics. *Adaptive Behavior* 7:5–16. [MW]
Clark, H. H. (1974) Semantics and comprehension. In: *Current trends in linguistics, vol. 12,* ed. T. Sebock. Mouton. [DAS]
Clarke, F. R. (1957) Constant-ratio rule for confusion matrices in speech communication. *Journal of the Acoustical Society of America* 29:715–20. [DWM]
Cohen, J. (1964) Dependency of the spectral reflectance curves of Munsell color chips. *Psychonomic Science* 1:369–70. [NB]
Cohen, N. J. & Eichenbaum, H. (1993) *Memory, amnesia, and the hippocampal system.* MIT Press. [BCL, rJBT]
Cooper, L. A. (1975) Mental rotation of random two-dimensional shapes. *Cognitive Psychology* 7:20–43. [arRNS]
(1976) Demonstrations of a mental analog of an external rotation. *Perception and Psychophysics* 19:296–302. [arRNS]
Cooper, L. A. & Shepard, R. N. (1973) Chronometric studies of the rotation of mental images. In: *Visual information processing,* ed. W. G. Chase. Academic Press. [rRNS]
(1984) Turning something over in the mind. *Scientific American* 251:106–14. [aRNS]
Corbin, H. H. (1942) The perception of grouping and apparent movement in visual depth. *Archives of Psychology,* No. 273. [aRNS]
Cortese, J. M. & Dyre, B. P. (1996) Perceptual similarity of shapes generated from Fourier descriptors. *Journal of Experimental Psychology: Human Perception and Performance* 22:133–43. [SE]
Cosmides, L. & Tooby, J. (1997) Dissecting the computational architecture of social inference mechanisms. In: *Characterizing human psychological adaptations,* Ciba Foundation Symposium 208. Wiley. [rRNS]
Cowie, F. (1998) *What's within? Nativism reconsidered.* Oxford University Press. [LTM]
Craik, K. J. W. (1943) *The nature of explanation.* Cambridge University Press. [aHB, rRNS]
Craven, B. J. & Foster, D. H. (1992) An operational approach to colour constancy. *Vision Research* 32:1359–66. [DHF]
Crist, E. (1999) *Images of animals: Anthropomorphism and animal mind.* Temple University Press. [aMK]
Cummins, R. (1986) Inexplicit representation. In: *The representation of knowledge and belief,* ed. M. Brand & R. Harnish. University of Arizona Press. [GO]
Cunningham, J. P. & Shepard, R. N. (1974) Monotone mapping of similarities into a general metric space. *Journal of Mathematical Psychology* 11:335–63. [rRNS]
Curtis, D. W., Paulos, M. A. & Rule, S. J. (1973) Relation between disjunctive reaction time and stimulus difference. *Journal of Experimental Psychology* 99:167–73. [aRNS]
Cutting, J. E. & Proffitt, D. R. (1982) The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology* 14:211–46. [GV]
Cutzu, F. & Edelman, S. (1996) Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Sciences USA* 93:12046–50. [SE]
Czeisler, C. A., Duffy, J. F., Shanahan, T. L., Brown, E. N., Mitchell, J. F., Rimmer, D. W., Ronda, J. M., Silva, E. J., Allan, J. S., Emens, J. S., Dijk, D.-J. & Kronauer, R. E. (1999) Stability, precision, and near-24-hour period of the human circadian pacemaker. *Science* 284:2177–81. [rHH]
Czerlinski, J., Gigerenzer, G. & Goldstein, D. G. (1999) How good are simple heuristics? In: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart.* Oxford University Press. [PMT]
Daffertshofer, A., van den Berg, C. & Beek, F. J. (1999) A dynamical model for mirror movements. *Physica D* 132:243–66. http://www.elsevier.nl/IVP/01672789/132/243/abstract.html [TDF]
Damasio, A. R. (1999) *The feeling of what happens.* Harcourt. [WCH]
Dawkins, R. (1999) *The extended phenotype: The long reach of the gene.* Oxford University Press. [rRNS]
Deacon, T. (1997) *The symbolic species: The co-evolution of language and the human brain.* Penguin. [JP]
Decock, L. (2001) The metrical structure of colour spaces. In: *Theories, technologies, instrumentalities of colour,* ed. B. Saunders. University Press of America. (forthcoming). [LD]
Dennett, D. C. (1982) Styles of mental representation. *Proceedings of the Aristotelian Society New Series* 83:213–26. [GO]
(1991) *Consciousness explained.* Little, Brown. [IG]
(1996) *Darwin's dangerous idea: Evolution and the meanings of life.* Touchstone Books. [FLB]
Dent-Read, C. & Zukow-Golding, P., eds. (1997) *Evolving explanations of development: Ecological approaches to organism-environment systems.* American Psychological Association. [JP]
Depew, D. (2000) The Baldwin effect: An archaeology. *Cybernetics and Human Knowing* 7(1):7–20. [JP]
Deutsch, G., Bourbon, W. T., Papanicolaou, A. C. & Eisenberg, H. M. (1988) Visuospatial tasks compared via activation of regional cerebral blood-flow. *Neuropsychologia* 26(3):445–52. [DHF]
DeValois, R. L. & DeValois, K. K. (1993) A multistage color model. *Vision Research* 33:1053–65. [KHP]
Dinse, H. (1990) A temporal structure of cortical information processing. *Concepts in Neuroscience* 1:199–238. [AR]
Ditzinger, T. & Haken, H. (1989) Oscillations in the perception of ambiguous patterns. *Biological Cybernetics* 61:279–87. [TDF]
(1990) The impact of fluctuations on the recognition of ambiguous patterns. *Biological Cybernetics* 63:453–56. [TDF]
Dresp, B. (1993) Bright lines and edges facilitate the detection of small light targets. *Spatial Vision* 7:213–25. [BD]
(1999) Dynamic characteristics of spatial mechanisms coding contour structures. *Spatial Vision* 12:129–42. [BD]
Driver, P. M. & Humphries, D. A. (1988) *Protean behavior: The biology of unpredictability.* Oxford University Press. [PMT]
Duda, R. O., Hart, P. E. & Stork, D. G. (2000) *Pattern classification.* Wiley. [rJBT]
Duncker, K. (1929) Über induzierte Bewegung. *Psychologische Forschung* 12:180–259. [DT]
Durbin, J. R. (1985) *Modern algebra: An introduction, 2nd edition.* Wiley. [aMK]
Earman, J. (1992) *Bayes or bust? A critical examination of Bayesian confirmation theory.* MIT Press. [EH]
Edelman, G. (1989) *Neural Darwinism: The theory of neuronal group selection.* Basic Books. [KHP]
Edelman, G. & Tononi, G. (2000) *Consciousness: How matter becomes imagination.* Penguin. [JP]
Edelman, S. (1997) Computational theories of object recognition. *Trends in Cognitive Sciences* 1:296–304. [aMK]
(1999) *Representation and recognition in vision.* MIT Press. [SE, rRNS, AW]
Edelman, S., Grill-Spector, K., Kushnir, T. & Malach, R. (1999) Towards direct visualization of the internal shape representation space by fMRI. *Psychobiology* 26:309–21. [SE]
Edelman, S. & Intrator, N. (2000) (Coarse coding of shape fragments) + (Retinotopy) = Representation of structure. *Spatial Vision.* (in press). [SE]
Einstein, A. (1949) The problem of space, ether, and the field of physics. In: *Albert Einstein: Philosopher-scientist,* ed. P. A. Schilpp. The Library of Living Philosophers. [rRNS]
Eisenhart, L. P. (1961) *Continuous groups of transformations.* Dover. [WCH]
Ekman, G. (1954) Dimensions of color vision. *Journal of Psychology* 38:467–74. [aRNS]
Elder, J. H. & Goldberg, R. M. (1998) The statistics of natural image contours. In: *Proceedings of the IEEE Workshop on Perceptual Organisation in Computer Vision, 1998.* [aHB]
Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking innateness: A connectionist perspective on development.* MIT Press. [JP]
Emlen, S. T. (1975) The stellar-orientation system of a migratory bird. *Scientific American* 233:102–11. [KC]
Enright, J. T. (1972) A virtuoso isopod: Circa-lunar rhythms and their tidal fine structure. *Journal of Comparative Physiology* 77:141–62. [aRNS]
Epstein, W. (1994) Why do things look as they do?: What Koffka might have said to Gibson, Marr and Rock. In: *Gestalt psychology: It's origins, foundations and influence,* ed. S. Poggi. Leo S. Oschki Editore. [rRS]
Erickson, M. A. & Kruschke, J. K. (1998) Rules and exemplars in category learning. *Journal of Experimental Psychology: General* 127:107–40. [BCL]
Ermentout, G. B. & Cowan, J. D. (1979) A mathematical theory of visual hallucination patterns. *Biological Cybernetics* 34:137–50. [TDF]
Farrell, J. E. (1983) Visual transformations underlying apparent movement. *Perception and Psychophysics* 33:85–92. [LMP, aRNS, aDT]
Farrell, J. E. & Shepard, R. N. (1981) Shape, orientation, and apparent rotational motion. *Journal of Experimental Psychology: Human Perception and Performance* 7:477–86. [TDF, arRNS]
Feldman, J. (1997) The structure of perceptual categories. *Journal of Mathematical Psychology* 41:145–70. [arJBT, DV]
(2000) Minimization of Boolean complexity in human concept learning. *Nature* 407:630–33. [BCL]
Feyerabend, P. K. (1975) *Against method: Outline of an anarchistic theory of knowledge.* Humanities Press. [rHH]
Feynman, R. P. (1985) *QED: The strange theory of light and matter.* Princeton University Press. [rRNS]
Feynman, R. P. & Hibbs, A. R. (1965) *Quantum mechanics and path integrals.* McGraw-Hill. [rRNS]
Field, D. J. (1987) Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A* 4:2379–94. [aHB]
(1994) What is the goal of sensory coding? *Neural Computation* 6:559–601. [aHB]



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Field, D. J., Hayes, A. & Hess, R. F. (1993) Contour integration by the human visual system: Evidence for a local association field. *Vision Research* 33:173–93. [BD]

Fink, R. A. & Shepard, R. N. (1986) Visual functions of mental imagery. In: *Handbook of perception and human performance, vol. 1,* ed. K. R. Boff, L. Kauffman & J. P. Thomas. Wiley. [aMK]

Finlayson, G. D., Hordley, S. D. & Brill, M. H. (2000) Illuminant-invariance at a pixel. *Proceedings of the Eighth IS&T/SID Color Imaging Conference: Color Science and Engineering – Systems, Technologies, Applications,* pp. 85–90. IS&T: The Society for Imaging Science and Technology. [MHB]

Fischer, S., Kopp, C. & Dresp, B. (2000) A neural network model for long-range contour diffusion in vision. *Lecture Notes in Computer Science* 1811:336–42. [BD]

Fodor, J. (2000) *The mind doesn't work that way.* MIT Press. [SE, RM]

Fomenti, D. (2000) http://www.unipv.it/webbio/dfpaleoa.htm [WCH]

Forkman, B. & Vallortiga, G. (1999) Minimization of modal contours: An essential cross-species strategy in disambiguating relative depth. *Animal Cognition* 2:181–85. [GV]

Foster, D. H. (1973) An experimental examination of a hypothesis connecting visual pattern recognition and apparent motion. *Kybernetik* 14:63–70. [DHF]

(1975a) An approach to the analysis of the underlying structure of visual space using a generalized notion of visual pattern recognition. *Biological Cybernetics* 17:77–79. [DHF]

(1975b) Visual apparent motion of some preferred paths in the rotation group *SO*(3). *Biological Cybernetics* 18:81–89. [DHF, aHH, LMP, arRNS, aDT]

(1978) Visual apparent motion and the calculus of variations. In: *Formal theories of visual perception,* ed. E. L. J. Leeuwenberg & H. F. J. M. Buffart. Wiley. [DHF]

Foster, D. H., Amano, K. & Nascimento, S. M. C. (2001) How temporal cues can aid colour constancy. *Color Research and Application* 26 (Suppl.):S180–85. [DHF]

Foster, D. H. & Nascimento, S. M. C. (1994) Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London B* 257:115–21. [NB, DHF]

Frank, T. D., Daffertshofer, A., Beek, P. J. & Haken, H. (1999) Impacts of noise on a field theoretical model of the human brain. *Physica D* 127:233–49. [TDF]

Frank, T. D., Daffertshofer, A., Peper, C. E., Beek, P. J. & Haken, H. (2000) Towards a comprehensive theory of brain activity: Coupled oscillator systems under external forces. *Physica D* 144:62–86. http://www.elsevier.nl/IVP/01672789/127/233/abstract.html [TDF]

French, A. P. & Kennedy, P. J., eds. (1985) *Niels Bohr: A centenary volume.* Harvard University Press. [rRNS]

Freyd, J. J. (1983) Dynamic mental representations. *Psychology Review* 94:427–38. [aRNS]

Freyd, J. J. & Jones, K. T. (1994) Representational momentum for a spiral path. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 94:427–38. [aRNS]

Fried, L. S. & Holyoak, K. J. (1984) Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory and Cognition* 10:234–57. [RB, arJBT]

Frye, D., Zelazo, P. D. & Palfai, T. (1995) Theory of mind and rule-based learning. *Cognitive Development* 10:483–527. [BH]

Fukushima, K. (1980) Neocognition: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36:193–202. [TDF]

Gallant, J. L., Braun, J. & Van Essen, D. C. (1993) Selectivity for polar, hyperbolic, and Cartesian gratings in macaque visual cortex. *Science* 259:100–103. [rRNS]

Gallistel, C. R. (1990) *The organization of learning.* MIT Press. [SE, HK]

Gamkredlidze, R. V., ed. (1991) *Geometry I.* Springer-Verlag. [WCH]

Gärdenfors, P. (2001) *Conceptual spaces: The geometry of thought.* MIT Press. [SE]

Gardner-Medwin, A. R. & Barlow, H. B. (2001) The limits of counting accuracy distributed neural representations. *Neural Computation* 13:477–540. [aHB]

Garner, W. R. (1974) *The processing of information and structure.* Erlbaum. [aRNS]

Gelman, R., Durgin, F. & Kaufman, L. (1995) Distinguishing between animates and inanimates: Not by motion alone. In: *Causal cognition: A multidisciplinary debate,* ed. D. Sperber, D. Premack & A. J. Premack. Clarendon Press. [aHH]

Gentner, D. (1983) Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7:155–70. [BCL, rJBT]

Gentner, D. & Markman, A. B. (1997) Structure mapping in analogy and similarity. *American Psychologist* 52:45–56. [DG]

Gentner, D., Rattermann, M. J. & Forbus, K. D. (1993) The roles of similarity in transfer: Separating retrievability from inferential soundness. *Cognitive Psychology* 25:524–75. [DG, rJBT]

Georgopoulos, A., Lurito, J. T., Petrides, M., Schwartz, A. B. & Massey, J. T. (1988) Mental rotation of the neuronal population vector. *Science* 243:234–36. [SE, rRNS]

Gerbino, W. (1997) Figural completion. In: *Biocybernetics of vision: Integrative mechanisms and cognitive processes,* ed. C. Taddei-Ferretti. World Scientific. [WG]

Gewirth, A. (1982) *Human rights: Essays on justifications and applications.* University of Chicago Press. [rRNS]

Ghirlanda, S. & Enquist, M. (1998) Artificial neural networks as models of stimulus control. *Animal Behaviour* 56:1383–89. [KC]

(1999) The geometry of stimulus control. *Animal Behaviour* 58:695–706. [KC]

Gibson, E. J. & Walk, R. D. (1960) The "visual cliff." *Scientific American* 202:67–71. [FLB]

Gibson, J. J. (1950) *The perception of the visual world.* Houghton Mifflin. [DMJ, aMK]

(1966) *The senses considered as perceptual systems.* Houghton Mifflin. [DMJ, aMK]

(1979) *The ecological approach to visual perception.* Houghton Mifflin/Erlbaum. [arHH, DMJ, MKK, aMK, KKN, aRNS]

Gigerenzer, G. (1991) From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review* 98:254–67. [PMT]

Gigerenzer, G. & Goldstein, D. G. (1996) Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review* 103:650–69. [PMT]

Gigerenzer, G. & Todd, P. M. (1999a) Fast and frugal heuristics: The adaptive toolbox. In: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart.* Oxford University Press. [PMT]

(1999b) *Simple heuristics that make us smart.* Oxford University Press. [MDL, rRNS]

Gilbert, C. D. & Wiesel, T. N. (1990) The influence of contextual stimuli on the orientation selectivity of cells in the primary visual cortex of the cat. *Vision Research* 30:1689–701. [BD]

Gilbert, D. T. (1991) How mental systems believe. *American Psychologist* 46:107–19. [DAS]

Gleitman, L., Gleitman, H., Miller, C. A. & Ostrin, R. (1996) Similar, and similar concepts. *Cognition* 58:321–76. [rJBT]

Gluck, M. A. (1991) Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science* 2:50–55. [KC, arRNS, aJBT]

Goebel, P. (1990) The mathematics of mental rotations (Theoretical note). *Journal of Mathematical Psychology* 34:435–44. [rRNS]

Goethe, J. W. v. (1795) *Einleitung in die vergleichende Anatomie.* Böhlau. [AH]

Gogel, W. C. (1972) Scalar perceptions with binocular cues of distance. *American Journal of Psychology* 85:477–97. [FLB]

Goldstone, R. L. (1994) The role of similarity in categorization: Providing a groundwork. *Cognition* 52(2):125–57. [EMP, aJBT]

Goldstone, R. L., Gentner, D. & Medin, D. (1989) Relations relating relations. In: *Proceedings of the 11th Annual Meeting of the Cognitive Science Society,* pp. 131–38.. Erlbaum. [aJBT]

Goldstone, R. L., Lippa, Y. & Shffrin, R. M. (2001) Altering object representations through category learning. *Cognition* 78(1):45–88. [MDL]

Goldstone, R. L., Medin, D. L. & Gentner, D. (1991) Relational similarity and the non-independence of features in similarity judgments. *Cognitive Psychology* 23:222–64. [DG, BCL, rJBT]

Goodman, N. (1955) *Fact, fiction, and forecast.* Harvard University Press. [rRNS, arJBT]

(1972) Seven strictures on similarity. In: N. Goodman, *Problems and projects.* Bobbs-Merrill. [EMP, aJBT]

(1976) *Languages of art.* Hackett. [rRS]

(1990) Pictures in the mind? In: *Images and understanding: Thoughts about images; ideas about understanding,* ed. H. Barlow, C. Blakemore & M. Weston-Smith. Cambridge University Press. [aMK]

Green, B. F., Jr. (1961) Figure coherence in the kinetic depth effect. *Journal of Experimental Psychology* 62:272–82. [aRNS]

Griffiths, T. L. & Tenenbaum, J. B. (2000) Teacakes, trains, taxicabs, and toxins: A Bayesian account of predicting the future. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society,* pp. 202–207, ed. L. R. Gleitman & A. K. Joshi. Erlbaum. [MDL, arJBT]

Grossberg, S. (1999) How does the cerebral cortex work? Learning, attention, and grouping by the laminar circuits of visual cortex. *Spatial Vision* 12:163–85. [BD]

Grossman, M. & Wilson, M. (1987) Stimulus categorization by brain injured patients. *Brain and Cognition* 6:55–71. [KHP]

Grush, R. (1995) Emulation and cognition. Unpublished doctoral dissertation, University of California, San Diego. [MW]

Guggenheimer, H. W. (1977) *Differential geometry.* Dover. [WCH]

Güntürkün, O. (1996) Sensory physiology: Vision. In: *Sturkie's avian physiology,* ed. G. C. Whittow. Academic Press. [GV]

Guttman, N. & Kalish, H. I. (1956) Discriminability and stimulus generalization. *Journal of Experimental Psychology* 51:79–88. [DWM, arRNS]



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Hadamard, J. (1945) *An essay on the psychology of invention in the mathematical field.* Dover. [rRNS]
Hahn, U. & Chater, N. (1997) Concepts and similarity. In: *Knowledge, concepts, and categories,* ed. K. Lamberts & D. Shanks. Psychology Press/MIT Press. [EMP]
Hahn, U., Chater, N. & Richardson, L. B. (submitted) Similarity as transformation. [NC]
Haken, H. (1977) *Synergetics. An introduction.* Springer-Verlag. [TDF]
(1985) *Light II – Laser light dynamics.* North Holland. [TDF]
(1988) *Information and self-organization.* Springer. [AR]
(1991) *Synergetic computers and cognition.* Springer-Verlag. [TDF]
(1996) *Principles of brain functioning.* Springer-Verlag. [TDF]
Haken, H. & Stadler, M. (1990) *Synergetics of cognition.* Springer-Verlag. [TDF]
Hastings, M. H. (1997) Central clocking. *Trends in Neurosciences* 20:459–64. [MKn]
Hatfield, G. C. & Epstein, W. (1985) The status of the minimum principle in the theoretical analysis of visual perception. *Psychological Bulletin* 97:155–86. [WG]
Haussler, D. (1988) Quantifying inductive bias: AI learning algorithms and Valiant's learning framework. *Artificial Intelligence* 36:177–221. [rRNS]
Haussler, D., Kearns, M. & Schapire, R. E. (1994) Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension. *Machine Learning* 14:83:113. [arJBT]
Hearst, E. (1991) Psychology and nothing. *American Scientist* 79:432–43. [DAS]
Hecht, H. & Bertamini, M. (2000) Understanding projectile acceleration. *Journal of Experimental Psychology: Human Perception and Performance* 26:730–46. [aHH]
Hecht, H., Kaiser, M. K. & Banks, M. S. (1996) Gravitational acceleration as a cue for absolute size and distance? *Perception and Psychophysics* 58:1066–75. [aHH]
Hecht, H. & Proffitt, D. R. (1991) Apparent extended body motions in depth. *Journal of Experimental Psychology: Human Perception and Performance* 17:1090–103. [DHF, arHH, aDT]
(1995) The price of expertise: Effects of experience on the water-level task. *Psychological Science* 6:90–95. [aHH]
Heidegger, M. (1962) *Being and time.* Harper & Row. [MKn]
Heil, J. (1983) *Perception and cognition.* University of California Press. [JH]
Heit, E. (1997a) Features of similarity and category-based induction. In: *SimCat97: Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation,* ed. M. Ramscar, U. Hahn, E. Cambouropolos & H. Pain. Department of Artificial Intelligence, Edinburgh University. [aJBT]
(1997b) Knowledge and concept learning. In: *Knowledge, concepts, and categories,* ed. K. Lamberts & D. Shanks. Psychology Press. [EH, EMP]
(1998) A Bayesian analysis of some forms of inductive reasoning. In: *Rational models of cognition,* ed. M. Oaksford & N. Chater. Oxford University Press. [EH, arJBT]
(2000) Properties of inductive reasoning. *Psychonomic Bulletin and Review* 7:569–92. [EH]
Heit, E. & Bott, L. (2000) Knowledge selection in category learning. In: *Psychology of learning and motivation, vol. 39,* ed. D. L. Medin. Academic Press. [EH]
Hempel, C. G. (1965) *Aspects of scientific explanation.* Free Press. [rRS, rJBT]
(1966) *Philosophy of natural science.* Prentice-Hall. [WCH]
Hering, E. (1887/1964) *Outlines of a theory of the light sense.* Trans. from the German by L. M. Hurvich & D. Jameson, 1964. Harvard University Press. [aRNS]
Hertwig, R., Hoffrage, U. & Martignon, L. (1999) Quick estimation: Letting the environment do the work. In: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart.* Oxford University Press. [PMT]
Hildreth, E. C. & Koch, C. (1987) The analysis of visual motion: From computational theory to neural mechanisms. *Annual Review of Neuroscience* 10:477–533. [RM]
Hinton, G. & Nowlan, S. (1996) How learning can guide evolution. In: *Adaptive individuals in evolving populations,* ed. R. Belew & M. Mitchell. Addison-Wesley. [JP]
Hobbes, T. (1651) *Leviathon.* Printed for Andrew Crooke at the Green Dragon in St. Paul's Churchyard. (Original edition). [rRNS]
Hochberg, J. (1986) Representation of motion and space in video and cinematic displays. In: *Handbook of perception and human performance, vol. 1,* ed. K. J. Boff, L. Kaufman & J. P. Thomas. Wiley. [HI]
Hock, H. S., Kelso, J. A. S. & Schöner, G. (1993) Bistability and hysteresis in the organization of apparent motion patterns. *Journal of Experimental Psychology: Human Perception and Performance* 19:63–80. [TDF]
Hoffman, D. D. (1998) *Visual intelligence: How we create what we see.* Norton. [aHH]
Hoffman, D. D. & Bennett, B. M. (1986) The computation of structure from fixed-axiz motion: Rigid structures. *Biological Cybernetics* 54:71–83. [MB]
Hoffman, W. C. (1966) The Lie algebra of visual perception. *Journal of Mathematical Psychology* 3:65–98. [WCH, KHP]
(1978) The Lie transformation group approach to visual neuropsychology. In: *Formal theories of visual perception,* ed. E. L. J. Leeuwenberg & H. Buffart. Wiley. [WCH, rRNS]
(1989) The visual cortex is a contact bundle. *Applied Mathematics and Computation* 32:137–67. [WCH]
(1994) Conformal structures in perceptual psychology. *Spatial Vision* 8:19–31. [WCH]
(1997) Mind and the geometry of systems. In: *Two sciences of mind: Readings in cognitive science and consciousness,* ed. S. O. O'Nuallain, P. McKevitt & E. Mac Aogain. John Benjamins. [WCH]
(1998) ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1998.volume.9/psyc.98.9.03.part-whole-perception.4.hoffman [WCH]
(1999) Dialectic – a universal for consciousness? *New Ideas in Philosophy* 17:251–69. [WCH]
Homa, D., Sterling, S. & Trepel, L. (1981) Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory* 7:418–39. [EH]
Hood, B. M. (1995) Gravity rules for 2- to 4-year-olds? *Cognitive Development* 10:577–98. [BH, HK]
(1998) Gravity does rule for falling events. *Developmental Science* 1:59–63. [BH]
Hood, B. M., Hauser, M., Anderson, L. & Santos, L. (1999) Gravity biases in a non-human primate. *Developmental Science* 2:35–41. [BH]
Hood, B. M., Santos, L. & Fieselman, S. (2000) Two-year-olds' naive predictions for horizontal trajectories. *Developmental Science* 3:328–32. [BH, HK]
Horn, L. R. (1989) *A natural history of negation.* University of Chicago Press. [DAS]
Horwich, P. (1982) *Probability and evidence.* Cambridge University Press. [EH, rJBT]
Howard, I. (1978) Recognition and knowledge of the water-level principle. *Perception* 7:151–60. [aHH]
Howson, C. & Urbach, P. (1993) *Scientific reasoning: The Bayesian approach.* Open Court. [rJBT]
Hoyt, D. F. & Taylor, C. R. (1981) Gait and energetics of locomotion in horses. *Nature* 292:239–40. [TDF]
Hubel, D. H. & Wiesel, T. N. (1959) Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology* 148:574–91. [aHB]
Hull, D. (1976) Are species really individuals? *Systematic Zoology* 25:174–91. [AH]
Hume, D. (1739/1978) *A treatise of human nature.* Oxford University Press. [rJBT]
Hummel, J. E. (2000) Where view-based theories of human object recognition break down: The role of structure in human shape perception. In: *Cognitive dynamics: Conceptual change in humans and machines,* ed. E. Dietrich & A. Markman. Erlbaum. [SE]
Hurvich, L. M. (1981) *Color vision.* Sinauer. [VW]
Hurvich, L. M. & Jameson, D. (1957) An opponent-process theory of color vision. *Psychological Review* 64:384–404. [aRNS]
Hyvarinen, A. & Oja, E. (1996) Simple neuron models for independent component analysis. *International Journal for Neural Systems* 7:671–87. [aHB]
Indow, T. (1988) Multidimensional studies of Munsell color solid. *Psychological Review* 95:456–70. [LD]
(1999) Global structure of visual space as a united entity. *Mathematical Social Sciences* 38(3):377–92. [DHF]
Ingold, T. (1996) A comment on the distinction between the material and the social. *Ecological Psychology* 8(2):183–87. [JP]
Intraub, H., Bender, R. S. & Mangels, J. A. (1992) Looking at pictures but remembering scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:180–91. [HI]
Intraub, H. & Bodamer, J. L. (1993) Boundary extension: Fundamental aspect of pictorial representation or encoding artifact? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:1387–97. [HI]
Intraub, H., Gottesman, C. V. & Bills, A. (1998) Effect of perceiving and imagining scenes on memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:186–201. [HI]
Intraub, H., Gottesman, C. V., Willey, E. V. & Zuk, I. J. (1996) Boundary extension for briefly glimpsed photographs: Do common perceptual processes result in unexpected memory distortions? *Journal of Memory and Language* 35:118–34. [HI]
Intraub, H. & Richardson, M. (1989) Wide-angle memories of close-up scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:179–87. [HT]
Irwin, D. E. (1991) Perceiving an integrated visual world. In: *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience,* ed. D. E. Meyer & S. Kornblum. MIT Press. [HI]



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Ittelson, W. (1960) *Visual space perception.* Springer. [aMK]
Izmailov, C. A. & Sokolov, E. N. (1991) Spherical model of color and brightness discrimination. *Psychological Science* 2:244–59. [ENS]
(2000) Psychophysics beyond sensation: Subjective and objective scaling of large color differences. (in press). [ENS]
Jackendoff, R. (1987) *Consciousness and the computational mind.* MIT Press. [LTM]
Jacobs, D. M., Runeson, S. & Michaels, C. F. (in press) Learning to visually perceive the relative mass of colliding balls in globally and locally constrained task ecologies. *Journal of Experimental Psychology: Human Perception and Performance.* [DMJ]
Jacobs, G. H. (1981) *Comparative color vision.* Academic. [LTM]
(1992) Data and interpretation in comparative color vision. *Behavioral and Brain Sciences* 15:40–41. [LD]
(1993) The distribution and nature of colour vision among the mammals. *Biological Review* 68:413–71. [LTM]
(1996) Primate photopigments and primate color vision. *Proceedings of the National Academy of Sciences USA* 93:577–81. [IG]
Jacobs, G. H., Neitz, M., Deegan, J. F. & Neitz, J. (1996) Trichromatic colour vision in New World monkeys. *Nature* 382:156–58. [IG]
Jacoby, L. L. (1983) Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior* 22:484–508. [BCL]
James, W. (1890/1950) *The principles of psychology, vols. 1 and 2.* Holt/Dover. (Original work published in 1890). [aHH, RM, aRNS]
(1892) *Psychology: Briefer course.* Holt. (Abridged edition of 1890 *Principles of psychology*). [DAS]
Jaynes, E. T. (1978) Where do we stand on maximum entropy? In: *The maximum entropy formalism,* ed. R. D. Levine & M. Tribus. MIT Press. [aRNS]
Johansson, G. (1950) *Configurations in event perception.* Almqvist & Wiksell. [DT]
(1964) Perception of motion and changing form. *Scandinavian Journal of Psychology* 5:181–208. [aDT]
Johnson, H. M. & Seifert, C. M. (1994) Sources of the continued influence effect: When misinformation in memory affects later inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20(6):1420–36. [DAS]
Jones, C. D., Osorio, D. & Baddeley, R. J. (2001) Colour categorisation by domestic chicks. *Proceedings of the Royal Society of London B.* (submitted). [RB]
Jones, M. R. (1976) Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review* 83:323–55. [rRNS]
Jones, S. (1992) Natural selection in humans. In: *Human evolution,* ed. S. Jones, R. Martin & D. Pilbeam. Cambridge University Press. [AH]
Joseph, J. E. & Kubovy, M. (1994) Perception of patterns traced on the head. Poster presented at the Sixth Annual Meeting of the American Psychological Society, Washington, D. C., July 1994. [rRNS]
Judd, D. B., McAdam, D. L. & Wyszecki, G. (1964) Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America* 54:1031–40. [LD, aRNS]
Kaiser, M. K., Proffitt, D. R. & Anderson, K. A. (1985a) Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 11:795–803. [aHH, HK, rRNS]
Kaiser, M. K., Proffitt, D. R. & McCloskey, M. (1985b) The development of beliefs about falling objects. *Perception and Psychophysics* 38:533–39. [BH]
Kaiser, M. K., Proffitt, D. R., Whelan, S. & Hecht, H. (1992) The influence of animation on dynamical judgments: Informing all of the people some of the time. *Journal of Experimental Psychology: Human Perception and Performance* 18:669–90. [aHH, HK, rRNS]
Kanizsa, G. (1979) *Organization in vision.* Praeger. [GV]
Kanizsa, G. & Gerbino, W. (1982) Amodal completion: Seeing or thinking? In: *Organization and representation in perception,* ed. J. Beck. Erlbaum. [WG]
Kant, I. (1781/1968) *Critique of pure reason,* trans. N. K. Smith. St. Martins Press. [ACZ, rRNS]
(1785/1996) Groundwork to the metaphysics of morals. In: *Cambridge edition of the works of Immanuel Kant,* trans. & ed. M. Gregor. Cambridge University Press. [rRNS]
(1786/1970) *Metaphysical foundations of natural science,* trans. J. Ellington. Bobbs-Merrill. [ACZ]
Kapadia, M. K., Ito, M., Gilbert, C. D. & Westheimer, G. (1995) Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron* 15:843–56. [BD]
Kareev, Y. (2000) Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review* 107:397–402. [rJBT]
Karten, H. J. & Shimizu, T. (1989) The origins of neocortex: Connections and laminations as distinct events in evolution. *Journal of Cognitive Neuroscience* 1:291–301. [GV]
Kass, R. E. & Raftery, A. E. (1995) Bayes factors. *Journal of the American Statistical Association* 90(430):773–95. [MDL]
Kellman, P. J. & Shipley, T. F. (1991) A theory of visual interpolation in object perception. *Cognitive Psychology* 23:141–221. [WG]
Kelso, J. A. S. (1995) *Dynamic patterns – The self-organization of brain and behavior.* MIT Press. [TDF]
Kelso, J. A. S., Scholz, J. P. & Schöner, G. (1986) Non-equilibrium phase transitions in coordinated biological motion: Critical fluctuations. *Physics Letters A* 118:279–84. [TDF]
Kilpatrick, F. P. (1961) *Explorations in transactional psychology.* New York University Press. [aMK]
Kim, I. K. & Spelke, E. S. (1999) Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science* 2:339–62. [HK]
Kingdon, J. (1993) *Self-made man and his undoing.* Simon & Schuster. [JP]
Klein, F. (1893/1957) *Vorlesungen uber hochere geometrie (Lectures on higher geometry), 3rd edition.* Chelsea. (Original work published in 1893). [FLB, rRNS]
Knill, D. C. & Richards, W. A. (1996) *Perception as Bayesian inference.* Cambridge University Press. [aJBT, JRM]
Koffka, K. (1931) Die Wahrnehmung von Bewegung. In: *Handbook der normalen und pathologischen Physiologie, vol. 12, part 2,* ed. A. Bethe, et al. Springer-Verlag. [aRNS]
(1935) *Principles of Gestalt psychology.* Harcourt, Brace & World. [WG, arRNS, VW, ACZ]
Köhler, W. (1927) Zum Problem der Regulation. *Wilhelm Roux' Archiv für Entwicklungsmechanik der Organismen* 112:315–32. [ACZ]
(1929) *Gestalt psychology.* Liveright/New American Library. [WG,VW]
(1938) *The place of value in a world of facts.* Liveright. [rRNS]
(1958) The present situation in brain physiology. *American Psychologist* 13:150–54. [ACZ]
Kolers, P. A. (1972) *Aspects of motion perception.* Pergamon. [DHF, WG]
Kolers, P. A. & Pomerantz, J. R. (1971) Figural change in apparent motion. *Journal of Experimental Psychology* 87:99–108. [MB, aRNS]
Korte, A. (1915) Kinematoskopische Untersuchungen. *Zeitschrift für Psychologie* 72:193–296. [aRNS]
Kourtzi, Z. & Shiffrar, M. (1997) One-shot view invariance in a moving world. *Psychological Science* 8:461–66. [rRNS]
(1999a) Dynamic representations of human body movement. *Perception* 28:49–62. [rRNS]
(1999b) The visual representation of three-dimensional, rotating objects. *Acta Psychologica* 102:265–92. [rRNS]
Krantz, D. H. (1975a) Color measurement and color theory: I. Representation theorem for Grassman structures. *Journal of Mathematical Psychology* 12:283–303. [aRNS]
(1975b) Color measurement and color theory: II. Opponent-colors theory. *Journal of Mathematical Psychology* 12:204–327. [aRNS]
(1983) A comment on the development of the theory of "direct" psychophysics. *Journal of Mathematical Psychology* 27:325. [rRNS]
(1989) Color and force measurement. In: *Foundations of measurement, vol. II: Geometrical, threshold, and probabilistic representations,* ed. P. Suppes, D. Krantz, R. Luce & A. Tversky. Academic Press. [LD]
Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971) *Foundations of measurement, vol. 1: Additive and polynomial representations.* Academic Press. [aMK]
Krist, H. (2000) Action knowledge does not appear in any action context: Children's developing intuitions about trajectories. (submitted). [HK]
(2001) Development of naive beliefs about moving objects: The straight-down belief in action. *Cognitive Development* 15:397–424. [HK]
Krist, H., Fieberg, E. L. & Wilkening, F. (1993) Intuitive physics in action and judgment: The development of knowledge about projectile motion. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19:952–66. [aHH, HK]
Krist, H., Loskill, J. & Schwarz, S. (1996) Intuitive Physik in der Handlung: Perzeptiv-motorisches Wissen über Flugbahnen bei 5–7 jährigen Kindern [Intuitive physics in action: Perceptual-motor knowledge about trajectories in 5–7-year-old children]. *Zeitschrift für Psychologie* 204:339–66. [aHH, HK]
Krumhansl, C. L. & Kessler, F. J. (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review* 89:334–68. [aRNS]
Kruschke, J. K. (1992) ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99:22–44. [MDL, aJBT]
Kruse, P., Carmesin, H.-O., Pahlke, L., Strüber, D. & Stadler, M. (1996) Continuous phase transitions in the perception of multistable visual patterns. *Biological Cybernetics* 75:321–30. [TDF]
Kruskal, J. B. (1964a) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27. [arRNS]
(1964b) Nonmetric multidimensional scaling: A numerical method. *Psychometrika* 29:115–29. [rRNS]
Kubovy, M. (1983) Mental imagery majestically transforming cognitive psychology


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

(Review of R. N. Shepard & L. Cooper, *Mental images and their transformations*). *Contemporary Psychology* 28:661–63. [aMK, rRNS]
Kubovy, M. & Pomerantz, J. R. (1981) *Perceptual organization.* Erlbaum. [WG, ACZ]
Kubovy, M. & Wagemans, J. (1995) Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science* 6:225–34. [AR]
Kucera, H. & Francis, W. N. (1967) *Computational analysis of present-day American English.* Brown University Press. [rJBT]
Kurthen, M. (1992) *Neurosemantik.* Enke Verlag. [MKn]
Lackner, J. R. & Dizio, P. (1998) Gravitoinertial force background level affects adaptation to coriolis force perturbations of reaching movements. *Journal of Neurophysiology* 80:546–53. [aHH]
Lacquaniti, F. (1996) Neural control of limb mechanics for visuomanual coordination. In: *Hand and brain: The neurophysiology and psychology of hand function,* ed. A. M. Wing, P. Haggard & J. R. Flanagan. Academic Press. [FL]
Lacquaniti, F., Borghese, N. A. & Carrozzo, M. (1992) Internal models of limb geometry in the control of hand compliance. *Journal of Neuroscience* 12:1750–62. [FL]
Lacquaniti, F., Carrozzo, M. & Borghese, N. A. (1993a) The role of vision in tuning anticipatory motor responses of the limbs. In: *Multisensory control of movement,* ed. A. Berthoz. Oxford University Press. [FL]
(1993b) Time-varying mechanical behavior of multijointed arm in man. *Journal of Neurophysiology* 69:1443–64. [FL]
Lacquaniti, F. & Maioli, C. (1989a) The role of preparation in tuning anticipatory and reflex responses during catching. *Journal of Neuroscience* 9:134–48. [FL]
(1989b) Adaptation to suppression of visual information during catching. *Journal of Neuroscience* 9:149–59. [FL]
Lakatos, S. (1993) Recognition of complex auditory-spatial patterns. *Perception* 22:363–74. [rRNS]
Lakatos, S. & Shepard, R. N. (1997a) Constraints common to apparent motion in visual, tactile, and auditory space. *Journal of Experimental Psychology: Human Perception and Performance* 23:1050–60. [rRNS]
(1997b) Time-distance relations in shifting attention between locations on one's body. *Perception and Psychophysics* 59:557–66. [rRNS]
Lakoff, G. & Johnson, M. (1990) *Metaphors we live by.* University of Chicago Press. [arMK, MW]
(1999) *Philosophy in the flesh: The embodied mind and its challenge to western thought.* Basic Books. [aMK, MKn]
Land, E. H. (1986) Recent advances in retinex theory. *Vision Research* 26:7–21. [LD]
Land, E. H. & McCann, J. J. (1971) Lightness and retinex theory. *Journal of the Optical Society of America* 61:1–11. [aRNS]
Landauer, T. K. & Dumais, S. T. (1997) A solution to Plato's problem: The Latent Semantic Analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104:211–40. [SE, rJBT]
Lee, D. N. (1980) Visuo-motor coordination in space-time. In: *Tutorials in motor behavior,* ed. G. E. Stelmach & J. Requin. Elsevier. [FL]
Lee, D. N. & Thomson, J. A. (1982) Vision in action: The control of locomotion. In: *Analysis of visual behavior,* ed. D. J. Ingle, M. A. Goodale & R. J. W. Mansfield. MIT Press. [HK]
Lee, D. N., Young, D. S., Reddish, P. E., Lough, S. & Clayton, T. M. (1983) Visual timing in hitting an accelerating ball. *Quarterly Journal of Experimental Psychology* 35A:333–46. [FL]
Lee, M. D. (1999) On the complexity of multidimensional scaling, additive tree, and additive clustering representations. Paper presented at the Symposium on Model Complexity held at the 32nd Annual Meeting of the Society for Mathematical Psychology, Santa Cruz, CA, July/August 1999. http://www.psychology.adelaide.edu.au/members/external/michaellee/modelcomp2.pdf [MDL]
(2001) On the complexity of additive clustering models. *Journal of Mathematical Psychology* 45(1):131–48. [MDL]
(submitted) A simple method for generating additive clustering models with limited complexity. http://www.psychology.adelaide.edu.au/members/external/michaellee/ac/pdf [aJBT]
Lennie, P. (1998) Single units and visual cortical organization. *Perception* 27:889–935. [aHB]
Lettwin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. (1959) What the frog's eye tells the frog's brain. *Proceedings of the IRE* 47:1940–59. [ACZ]
Lewicki, M. L. (2000) Efficient coding of natural sounds. Unpublished manuscript. [JRM]
Leyton, M. (1989) Inferring causal history from shape. *Cognitive Science* 13:357–87. [rHH]
(1992) *Symmetry, causality, mind.* MIT Press/Bradford Books. [arRNS, DV]
Li, M. & Vitanyi, P. M. B. (1997) *An introduction to Kolmogorov complexity and its applications, 2nd edition.* Springer-Verlag. [NC]
Liben, L. S. (1991) The Piagetian water-level task: Looking beneath the surface. In: *Annals of child development 8,* ed. R. Vasta. Kingsley. [aHH]
Linsker, R. (1986a) From basic network principles to neural architecture – emergence of spatial opponent cells. *Proceedings of the National Academy of Sciences USA* 83:7508–12. [aHB]
(1986b) From basic network principles to neural architecture – emergence of orientation selective cells. *Proceedings of the National Academy of Sciences USA* 83:8390–94. [aHB]
(1986c) From basic network principles to neural architecture – emergence of orientation columns. *Proceedings of the National Academy of Sciences USA* 83:8779–83. [aHB]
Lockhead, G. R. (1966) Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology* 72:95–104. [aRNS]
Loomis, J. M. & Nakayama, K. (1973) A velocity analogue of brightness contrast. *Perception* 2:425–28. [DT]
Love, B. C., Markman, A. B. & Yamauchi, T. (2000) Modeling inference and classification learning. *The National Conference on Artificial Intelligence (AAAI–2000):*136–41. [BCL]
Love, B. C. & Medin, D. L. (1998) Sustain: A model of human category learning. In: *Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI–98),* pp. 671–76. AAAI Press/MIT Press. [arJBT]
Lubin, J. (1995) A visual system discrimination model for imaging system design and evaluation. In: *Visual models for target detection and recognition,* ed. E. Peli. World Scientific. [MHB]
Luce, R. D. (1959) *Individual choice behavior; a theoretical analysis.* Wiley. [DWM]
Luria, S. E. & Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28:491–511. [AH]
Lythgoe, J. N. (1979) *The ecology of vision.* Clarendon Press. [LTM]
Mach, E. (1886/1922) *Die Analyse der Empfindungen und das Verhältnis des Physischen zum Psychischen.* (First German edition, 1886). Gustav Fischer. [rHH] *The analysis of sensations, and the relation of the physical to psychical* (Translation of the 1st, revised from the 5th, German edition, by S. Waterlow). Open Court. (Dover reprint, 1959). [aHB]
Mack, A. & Herman, E. (1972) A new illusion: The underestimation of distance during pursuit eye movements. *Perception and Psychophysics* 12:471–73. [DT]
MacKay, D. J. C. (1992) Bayesian interpolation. *Neural Computation* 4:415–47. [aHB, rJBT]
Mackintosh, N. J. (1983) *Conditioning and associative learning.* Oxford University Press. [aHB]
MacLennan, B. (1999) Field computation in natural and artificial intelligence. *Information Sciences* 119:73–89. [SE]
Maddox, W. T. & Bohil, C. J. (1998) Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:1459–82. [EH]
Maloney, L. T. (1986) Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America A* 3:1673–83. [NB, aRNS]
(1999) Physics-based approaches to modeling surface color perception. In: *Color vision: From genes to perception,* ed. K. R. Gegenfurtner & L. T. Sharpe. Cambridge University Press. [DHF]
Maloney, L. T. & Wandell, B. A. (1986) Color constancy: A method for recovering surface spectral reflectance. *Journal of the Optical Society of America A* 3:29–33. [LD, arRNS]
Malt, B. C. (1994) Water is not $H_20$. *Cognitive Psychology* 27:41–70. [EMP]
Man, T. & MacAdam, D. (1989) Three-dimensional scaling of the uniform color scales of the Optical Society of America. *Journal of the Optical Society of America A* 6:128–38. [LD]
Mandelbrot, B. B. (1983) *The fractal geometry of nature.* W. H. Freeman. [DV]
Marchant, J. A. & Onyango, C. M. (2000) Shadow-invariant classification for scenes illuminated by daylight. *Journal of the Optical Society of America A* 17:1952–61. [MHB]
Marimont, D. H. & Wandell, B. A. (1992) Linear models of surface and illuminant spectra. *Journal of the Optical Society of America A* 9:1905–13. [aRNS]
Marr, D. (1970) A theory for cerebral neocortex. *Proceedings of the Royal Society of London B* 176:161–234. [SE]
(1982) *Vision: A computational investigation into the human representation and processing of visual information.* W. H. Freeman. [FLB, aMK, JRM, LMP, arJBT]
Martignon, L. & Hoffrage, U. (1999) Why does one-reason decision making work? A case study in ecological rationality. In: G. Gigerenzer, P. M. Todd & the ABC Research Group, *Simple heuristics that make us smart.* Oxford University Press. [rRNS, PMT]
Massaro, D. W. (1996) Integration of multiple sources of information in language processing. In: *Attention and performance XVI: Information integration in perception and communication,* ed. T. Inui & J. L. McClelland. MIT Press. [DWM]



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

(1998) *Perceiving talking faces: From speech perception to a behavioral principle.* MIT Press. [DWM]
Massaro, D. W. & Cohen, M. M. (2000) Tests of auditory-visual integration efficiency within the framework of the fuzzy logical model of perception. *Journal of the Acoustical Society of America* 108:784–89. [DWM]
Massaro, D. W. & Stork, D. G. (1998) Sensory integration and speechreading by humans and machines. *American Scientist* 86:236–44. [DWM]
Massey, C. M. & Gelman, R. (1988) Preschooler's ability to decide whether a photographed unfamiliar object can move itself. *Developmental Psychology* 24:307–17. [aHH]
Massironi, M. & Luccio, R. (1989) Organizational versus geometric factors in mental rotation and folding tasks. *Perception* 18:321–32. [JRP]
Maturana, H. R. & Varela, F. J. (1987) *The tree of knowledge.* New Science Library. [MKn]
Mausfeld, R. (in press) The physicalistic trap in perception theory. In: *Perception and the physical world,* ed. D. Heyer & R. Mausfeld. Wiley. [RM]
Mausfeld, W., Niederée, R. & Heyer, D. (1992) On possible perceptual worlds and how they shape their environment. *Behavioral and Brain Sciences* 15:47–48. [LD]
Maynard-Smith, J. (1968) *Mathematical ideas in biology.* Cambridge University Press. [rJBT]
Mayr, E. (1976) *Evolution and the diversity of life.* Harvard University Press. [JP]
McAfee, E. A. & Proffitt, D. R. (1991) Understanding the surface orientation of liquids. *Cognitive Psychology* 23:483–514. [aHH, HK]
McBeath, M. K., Morikawa, K. & Kaiser, M. (1992) Perceptual bias for forward-facing motion. *Psychological Science* 3:462–67. [MB]
McBeath, M. K., Schaffer, D. M. & Kaiser, M. K. (1995) How baseball outfielders determine where to run to catch fly balls. *Science* 268:569. [FL]
McBeath, M. K. & Shepard, R. N. (1989) Apparent motion between shapes differing in location and orientation: A window technique for estimating path curvature. *Perception and Psychophysics* 46(4):333–37. [MB, DHF, aHH, aRS, arRNS, aDT]
McCann, J. (1992) Rules for colour constancy. *Opthalmological and Physiological Optics* 12:175–77. [NB]
McCloskey, M. (1983) Intuitive physics. *Scientific American* 248:122–30. [arRNS]
McCloskey, M., Caramazza, A. & Green, B. (1980) Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science* 210:1139–41. [aHH, rRNS]
McCloskey, M. & Kohl, D. (1983) Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9:146–56. [HK]
McCloskey, M., Washburn, A. & Felch, L. (1983) Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 9:636–49. [aHH, BH, HK]
McConnell, D. S., Muchisky, M. M. & Bingham, G. P. (1998) The use of time and trajectory forms as visual information about spatial scale in events. *Perception and Psychophysics* 60(7):1175–87. [AW]
McIntyre, J., Zago, M., Berthoz, A. & Lacquaniti, F. (1999) Internal models for ball catching studied in 0g. *Society for Neuroscience Abstracts* 25:115. [FL, rRNS]
McLeod, P. & Dienes, Z. (1996) Do fielders know where to go to catch the ball or only how to get there? *Journal of Experimental Psychology: Human Perception and Performance* 22:531–43. [rRNS, PMT]
Medin, D. L., Goldstone, R. & Gentner, D. (1993) Respects for similarity. *Psychological Review* 100:254–78. [DG, aJBT]
Medin, D. L. & Ortony, A. (1989) Psychological essentialism. In: *Similarity and analogical reasoning,* ed. S. Vosniadou & A. Ortony. Cambridge University Press. [EMP]
Meltzoff, A. N. & Moore, M. K. (1977) Imitation of facial and manual gestures by human neonates. *Science* 198:75–78. [rRNS]
(1999) Resolving the debate about early imitation. In: *Reader in developmental psychology,* ed. A. Slater & D. Muir. Blackwell. [rRNS]
Menzel, R. & Backhaus, W. (1991) Colour vision in insects. In: *The perception of colour,* ed. P. Gouras. CRC Press. [IG]
Metzger, W. (1941/1954) *Psychologie.* Steinkopff. (1st edition, 1941). Spanish edition, 1954: *Psicologia,* trans. H. W. Jung. Editorial Nova. Italian edition, 1971: *I fondamenti della psicologia della gestalt,* trans. L. Lumbelli. Giunti-Barbèra. [WG]
(1975) *Gesetze des Sehens, 3rd edition.* Kramer. [aHH, aDT]
Metzler, J. & Shepard, R. N. (1974) Transformational studies of the internal representation of three-dimensional objects. In: *Theories of cognitive psychology: The Loyola symposium,* ed. R. Solso. Erlbaum. [arRNS]
Michaels, C. F. & de Vries, M. (1998) Higher-order and lower-order variables in the visual perception of relative pulling force. *Journal of Experimental Psychology: Human Perception and Performance* 24:526–46. [DMJ]
Michotte, A. (1963) *The perception of causality.* Basic Books. [GV]
Michotte, A., Thines, G. & Crabbe, G. (1964) *Les complements amodaux des structures perceptives* (*Amodal completion of perceptual structures*). *Studia Psychologica.* Publications Universitaires de Louvain. [GV]
Mićunović, M. & Kojić, M. (1988) *Kinematika.* Naučna Knjiga. [aDT]
Mill, J. S. (1843) *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence, and methods of scientific investigation.* J. W. Parker. [rJBT]
(1861/1863) *Utilitarianism.* (Originally published in Fraser's Magazine, London, 1861). [rRNS]
Miller, G. F. & Shepard, R. N. (1993) An objective criterion for apparent motion based on phase discrimination. *Journal of Experimental Psychology: Human Perception and Performance* 19:48–62. [aRNS]
Milner, A. D. & Goodale, M. A. (1995) *The visual brain in action.* Oxford University Press. [aMK, rRNS]
Minetti, A. E. & Alexander, R. M. (1997) A theory of metabolic costs for bipedal gaits. *Journal of Theoretical Biology* 29:195–98. [TDF]
Mitchell, T. M. (1990) The need for biases in learning generalization. In: *Readings in machine learning,* ed. J. W. Shavlik & T. G. Dietterich. Morgan Kaufman. (Originally published as a *Rutgers Technical Report,* 1980). [rRNS]
(1997) *Machine learning.* McGraw Hill. [arJBT]
(1999) The role of unlabeled data in supervised learning. In: *Proceedings of the Sixth International Colloquium on Cognitive Science.* http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-20/www/mitchell-pubs/iccs99.ps [aJBT]
Mollon, J. D. (1989) "Tho' she kneel'd in that place where they grew . . .": The uses and origins of primate colour vision. *Journal of Experimental Biology* 146:21–38. [NB]
(1995) Seeing colour. In: *Colour: Art and science,* ed. T. Lamb & J. Bourriau. Cambridge University Press. [rMK]
Mollon, J. D. & Jordan, G. (1993) A study of women heterozygous for colour deficiencies. *Vision Research* 33:1495–508. [LD]
Mori, T. (1982) Apparent motion path composed of a serial concatenation of translations and rotations. *Biological Cybernetics* 44:31–34. [LMP, aDT]
Movellan, J. R. & McClelland, J. L. (2001) The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review* 108(1):113–48. [JRM]
Muchisky, M. M. & Bingham, G. P. (in press) Trajectory forms as a source of information about events. *Perception and Psychophysics.* [AW]
Mumsford, D. (1994) Neuronal architectures for pattern-theoretic problems. In: *Large-scale neuronal theories of the brain,* ed. C. Koch & J. L. Davis. MIT Press. [SE]
(1996) Pattern theory: A unifying perspective. In: *Perception as Bayesian inference,* ed. D. C. Knill & W. Richards. Cambridge University Press. [aHB]
Mundy, J. L., Zisserman, A. & Forsyth, D., eds. (1994) *Applications of invariance in computer vision.* Springer-Verlag. [KKN]
Murphy, G. L. & Medin, D. L. (1985) The role of theories in conceptual coherence. *Psychological Review* 92:289–316. [EH, EMP]
Murphy, G. L. & Ross, B. H. (1994) Predictions from uncertain categorizations. *Cognition Psychology* 27:148–93. [BCL]
Myung, I. J. (1994) Maximum entropy interpretation of decision bound and context models of categorization. *Journal of Mathematical Psychology* 38:335–65. [aRNS]
Myung, I. J. & Pitt, M. A. (1997) Applying Occam's Razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin and Review* 3(2):227–30. [MDL]
Myung, I. J. & Shepard, R. N. (1996) Maximum entropy inference and stimulus generalization (Theoretical Note). *Journal of Mathematical Psychology* 40:342–47. [rRNS]
Nagle, M. G. & Osorio, D. (1993) The tuning of human photopigments may minimize red-green chromatic signals in natural conditions. *Proceedings of the Royal Society of London B* 252:209–13. [NB]
Nánez, J. E. (1988) Perception of impending collision in 3- to 6-week-old human infants. *Infant Behavior and Development* 11:447–63. [HK]
Nascimento, S. M. C. & Foster, D. H. (1997) Detecting natural changes of cone-excitation ratios in simple and complex coloured images. *Proceedings of the Royal Society of London B* 264:1395–402. [NB, DHF]
Neal, R. M. (2000) Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 9:249–65. [rJBT]
Neisser, U. (1982) *Memory observed: Remembering in natural contexts.* W. H. Freeman. [WCH]
Niall, K. K. (1997) "Mental rotation," pictured rotation, and tandem rotation in depth. *Acta Psychologica* 95(1):31–83. [KKN]
Nickerson, D. (1981) OSA uniform color scale samples: A unique set. *Color Research and Application* 6:7–33. [LD]
Nicolis, G. (1995) *Introduction to nonlinear sciences.* Cambridge University Press. [TDF]
Nikolaev, P. P. (1985) Model of constant pattern of colour perception for the case of continuous spectral functions. *Biofizika* 30:112–17. [MHB]


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Nisbett, R. & Ross, L. (1980) *Human inference: Strategies and shortcomings of social judgment.* Prentice-Hall. [DAS]
Noll, A. M. (1965) Computer-generated three-dimensional movies. *Computers and Animation* 14:20–23. [aRNS]
Nosofsky, R. M. (1986) Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115:39–57. [NC, BCL, rRNS, aJBT]
(1987) Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:87–108. [aRNS]
(1989) Further tests of an exemplar-similarity approach to relating identification and categorization. *Journal of Experimental Psychology: Perception and Psychophysics* 45:279–90. [EMP]
(1998a) Optimal performance and exemplar models of classification. In: *Rational models of cognition,* ed. M. Oaksford & N. Chater. Oxford University Press. [arJBT]
(1988b) Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14:54–65. [EH]
(1992) Similarity scaling and cognitive process models. *Annual Review of Psychology* 43:25–53. [aRNS]
Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C. & Glauthier, P. (1994) Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory and Cognition* 22:352–69. [aRNS]
Nosofsky, R. M., Kruschke, J. K. & McKinley, S. (1992) Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18:211–33. [aRNS]
Nuboer, J. (1986) A comparative view on colour vision. *Netherlands Journal of Zoology* 36:344–80. [LD]
Oaksford, M. & Chater, N. (1998) *Rational models of cognition.* Oxford University Press. [JRM, rJBT]
(1999) Ten years of the rational analysis of cognition. *Trends in Cognitive Science* 3:57–65. [rJBT]
O'Brien, G. & Opie, J. (1999) A connectionist theory of phenomenal experience. *Behavioral and Brain Sciences* 22:127–48. [GO]
(forthcoming) Notes toward a structuralist theory of mental representation. In: *Representation in mind: New approaches to mental representation,* ed. H. Clapin, P. Staines & P. Slezak. Greenwood. [GO]
Oden, G. C. (1977) Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance* 3:565–75. [DWM]
Oden, G. C. & Massaro, D. W. (1978) Integration of featural information in speech perception. *Psychological Review* 85:172–91. [DWM, rRNS]
Ogasawara, J. (1936) Effect of apparent separation on apparent movement. *Japanese Journal of Psychology* 11:109–22. [aRNS]
Olivetti Belardinelli, M. (1976) *La costruzione della realtà.* Boringhieri. [AR]
Oller, K. (2000) *The emergence of the speech capacity.* Erlbaum. [AH]
Olshausen, B. A. & Field, D. J. (1996) Emergence of simple-cell receptive-field properties by learning a sparse code for natural images. *Nature* 381:607–609. [aHB]
O'Regan, J. K. (1992) Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology* 46:461–88. [HI]
Osherson, D. N. & Smith, E. E. (1981) On the adequacy of prototype theory as a theory of concepts. *Cognition* 9:35–58. [DWM]
(1982) Gradedness and conceptual combination. *Cognition* 12:299–318. [DWM]
Osherson, D. N., Smith, E. E., Wilkie, O., Lopez, A. & Shafir, E. (1990) Category-based induction. *Psychological Review* 97(2):185–200. [EH, aJBT]
Osorio, D. & Bossomaier, T. R. (1992) Human cone-pigment spectral sensitivities and the reflectances of natural surfaces. *Biological Cybernetics* 67:217–22. [IG]
Osorio, D. & Vorobyev, M. (1996) Colour vision as an adaptation to frugivory in primates. *Proceedings of the Royal Society of London B* 263:593–99. [NB, IG]
Osorio, D., Vorobyev, M. & Jones, C. D. (1999) Colour vision of domestic chicks. *Journal of Experimental Biology* 202:2951–59. [IG]
Palmer, S. E. & Rock, I. (1994) Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin and Review* 1:29–55. [aHH]
Pani, J. R. (1989) Orientation of the object and slant of the axis of rotation in perception and mental imagery of object rotation. Poster presented at the Annual Meeting of the Psychonomic Society, Atlanta, GA, November 1989. [JRP]
(1993) Limits on the comprehension of rotational motion: Mental imagery of rotations with oblique components. *Perception* 22:785–808. [JRP, LMP]
(1997) Descriptions of orientation in physical reasoning. *Current Directions in Psychological Science* 6:121–26. [JRP]
(1999) Descriptions of orientation and structure in perception and physical reasoning. In: *Ecological approaches to cognition: Essays in honor of Ulric Neisser,* ed. E. Winograd, R. Fivush & W. Hirst. Erlbaum. [JRP]
Pani, J. R. & Dupree, D. (1994) Spatial reference systems in the comprehension of rotational motion. *Perception* 23:929–46. [JRP]
Pani, J. R., William, C. T. & Shippey, G. (1995) Determinants of the perception of rotational motion: Orientation of the motion to the object and to the environment. *Journal of Experimental Psychology: Human Perception and Performance* 21:1441–56. [JRP]
Parkinnen, J. P. S., Hallikainen, J. & Jaaskelainen, T. (1989) Characteristic spectra of surface Munsell colors. *Journal of the Optical Society of America A* 6:318–22. [NB]
Parsons, L. M. (1987a) Imagined spatial transformation of one's body. *Journal of Experimental Psychology: General* 116:172–91. [LMP]
(1987b) Imagined spatial transformation of one's hands and feet. *Cognitive Psychology* 19:178–241. [LMP]
(1987c) Visual discrimination of abstract mirror-reflected three-dimensional objects at any orientations. *Perception and Psychophysics* 42:29–59. [LMP]
(1995) Inability to reason about an object's orientation using an axis and angle of rotation. *Journal of Experimental Psychology: Human Perception and Performance* 21:1259–77. [JRP, LMP]
Pearson, K. (1892) *The grammar of science.* Walter Scott. [aHB]
Peirce, C. S. (1966) *Selected writings.* Dover. [DAS]
Peitgen, H.-O., Jürgens, H. & Saupe, D. (1992) *Chaos and fractals: New frontiers of science.* Springer-Verlag. [DV]
Peper, C. E., Beek, P. J. & van Wieringen, P. C. W. (1995) Multifrequency coordination in bimanual tapping: Asymmetrical coupling and signs of supercriticality. *Journal of Experimental Psychology: Human Perception and Performance* 21:1117–38. [TDF]
Petrov, A. P. (1992) Surface color and color constancy. *Color Research and Application* 18:236–40. [MHB]
Petter, G. (1956) Nuove ricerche sperimentali sulla totalizzazione percettiva. *Rivista di Psicologia* 50:213–27. [GV]
Piaget, J. (1954) *The construction of reality in the child.* Basic Books. [BH]
Pickering, M. & Chater, N. (1995) Why cognitive science is not formalized folk psychology. *Minds and Machines* 5:309–37. [EMP]
Pinker, S. (1984) *Language learnability and language learning.* MIT Press. [aMK]
Poggio, T. (1990) Vision: The 'other' face of AI. In: *Modeling the mind,* ed. K. A. M. Said, W. H. Newton-Smith, R. Viale & K. V. Wilkes. Clarendon Press. [aHH]
Poggio, T. & Shelton, C. (1999) Machine learning, machine vision and the brain. *AI Magazine* 3:37–55. [aJBT]
Poggio, T., Torre, V. & Koch, C. (1985) Computational vision and regularization theory. *Nature* 317:314–19. [aMK]
Popper, K. R. (1935) *Logik der Forschung: Zur Erkenntnistheorie der modernen Naturwissenschaft.* [*The logic of scientific discovery*]. J. Springer. [arHH]
(1972) *Objective knowledge: An evolutionary approach.* Clarendon Press. [rJBT]
(1978) Natural selection and the emergence of mind. *Dialectica* 32(3/4):339–55. [JP]
Port, R. F. & Van Gelder, T. (1995) *Mind as motion.* MIT Press. [DV]
Posner, M. I. & Keele, S. W. (1968) On the genesis of abstract ideas. *Journal of Experimental Psychology* 77:353–63. [rJBT]
Pothos, E. M. & Hahn, U. (2000) So concepts aren't definitions, but do they have necessary or sufficient features? *British Journal of Psychology* 91:439–50. [EMP]
Pribram, K. H. (1986) Convolution and matrix systems as content addressable distributed brain processes in perception and memory. *Journal of Neurolinguistics* 2(2):339–64. [KHP]
(1991) *Brain and perception.* Erlbaum. [KHP]
Pribram, K. H. & Carlton, E. H. (1987) Holonomic brain theory in imaging and object perception. *Acta Psychologica* 63:175–210. [KHP]
Prigogine, I. & Nicolis, G. (1987) *Exploring complexity. An introduction.* Piper. [AR]
Prinz, W. (1987) Ideo-motor action. In: *Perspectives on perception and action,* ed. H. Heuer & A. F. Sanders. Erlbaum. [aHH]
(1992) Why don't we perceive our brain states? *European Journal of Cognitive Psychology* 4:1–20. [rHH]
(1997) Perception and action planning. *European Journal of Cognitive Psychology* 9:129–54. [rHH]
Proffitt, D. R. & Gilden, D. L. (1989) Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance* 15:384–93. [arRNS]
Proffitt, D. R., Gilden, D. L., Kaiser, M. K. & Whelan, S. M. (1988) The effect of configural orientation on perceived trajectory in apparent motion. *Perception and Psychophysics* 43:465–74. [MB, aRNS, aDT]
Proffitt, D. R. & Kaiser, M. K. (1998) The internalization of perceptual processing



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

of constraints. In: *Perception and cognition at century's end: Handbook of perception and cognition, 2nd edition,* ed. J. Hochberg, E. C. Caterette & M. P. Friedman. Academic Press. [aHH, aMK]
Proffitt, D. R., Kaiser, M. K. & Whelan, S. M. (1990) Understanding wheel dynamics. *Cognitive Psychology* 22:342–73. [arRNS]
Prusinkiewicz, P. & Lindenmayer, A. (1990) *The algorithmic beauty of plants.* Springer-Verlag. [DV]
Pruzansky, S., Tversky, A. & Carroll, J. D. (1982) Spatial versus tree representations of proximity data. *Psychometrika* 47:3–24. [rRNS]
Putnam, H. (1980) *Reason, truth and history.* Cambridge University Press. [AR]
Pylyshyn, Z. W. (1984) *Computation and cognition.* MIT Press. [GO]
Quine, W. V. O. (1960) *Word and object.* MIT Press. [rJBT]
(1969) Natural kinds. In: W. V. Quine, *Ontological relativity, and other essays.* Columbia University Press. [aJBT]
Raffone, A. & Van Leeuwen, C. (in preparation) The graded synchrony hypothesis: Flexible coding in neural information processing. [AR]
Ramachandran, V. S. (1988a) The perception of depth from shading. *Scientific American* 269(2):76–83. [aHH, aDT]
(1988b) Perception of shape from shading. *Nature* 331:163–66. [FLB]
(1990a) Interactions between motion, depth, color and form: The utilitarian theory of perception. In: *Vision: Coding and efficiency,* ed. C. Blakemore. Cambridge University Press. [rRNS, PMT]
(1990b) Visual perception in people and machines. In: *AI and the eye,* ed. A. Blake & T. Troscianko. Wiley. [WG]
Ramscar, M. J. A. & Yarlett, D. G. (2000) A high-dimensional model of retrieval in analogy and similarity-based transfer. In: *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society,* ed. L. R. Gleitman & A. K. Joshi. Erlbaum. [rJBT]
Rappaport, R. A. (1999) *Ritual and religion in the making of humanity.* Cambridge University Press. [DAS]
Rawls, J. (1971) *A theory of justice.* Harvard University Press. [rRNS]
Reid, A. K. & Staddon, J. E. R. (1998) A dynamic route finder for the cognitive map. *Psychological Review* 105:585–601. [KC]
Rescorla, R. A. & Wagner, A. R. (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In: *Classical conditioning II: Current research and theory,* ed. A. H. Black & W. F. Prokasy. Appleton-Century-Crofts. [aHB]
Richard, A. F. (1992) Food in a primate's life. In: *Human evolution,* ed. S. Jones, R. Martin & D. Pilbeam. Cambridge University Press. [AH]
Richards, W. & Koenderink, J. J. (1995) Trajectory mapping (tm): A new non-metric scaling technique. *Perception* 24:1315–31. [rJBT]
Richter, W., Somorjai, R., Summers, R., Jarmasz, M., Menon, R. S., Gati, J. S., Georgopoulos, A. P., Tegeler, C., Ugurbil, K. & Kim, S.-G. (2000) Motor area activity during mental rotation studied by time-resolved single-trial fMRI. *Journal of Cognitive Neuroscience* 12(2):310–20. [DHF]
Riesenhuber, M. & Poggio, T. (2000) CBF: A new framework for object categorization in cortex. In: *Biologically motivated computer vision,* ed. S.-W. Lee, H. H. Bülthoff & T. Poggio. Springer-Verlag. [TDF]
Ringach, D. L., Hawken, M. J. & Shapley, R. (1997) Dynamics of orientation tuning in macaque primary visual cortex. *Nature (London)* 387:281–84. [aHB]
Rips, L. J. (1975) Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior* 14:665–81. [arJBT]
(1989) Similarity, typicality, and categorization. In: *Similarity and analogical reasoning,* ed. S. Vosniadou & A. Orton. Cambridge University Press. [arJBT]
Roberts, F. S. (1979/1984) *Measurement theory.* Cambridge University Press. (Original work published in 1979). [aMK]
Robin, D. J., Berthier, N. E. & Clifton, R. K. (1996) Infants' predictive reaching for moving objects in the dark. *Developmental Psychology* 32:824–35. [HK]
Robins, C. & Shepard, R. N. (1977) Spatio-temporal probing of apparent rotational movement. *Perception and Psychophysics* 22:12–18. [arRNS]
Rock, I. (1983) *The logic of perception.* MIT Press. [arHH, aMK, ACZ]
(1993) The logic of "The logic of perception." *Giornale Italiano di Psicologia* 20:841–67. [GV]
(1997) *Indirect perception.* MIT Press. [aMK]
Rock, I., Wheeler, D. & Tudor, L. (1989) Can we imagine how objects look from other viewpoints? *Cognitive Psychology* 21:185–210. [aRNS]
Roediger, H. L. & McDermott, K. B. (1995) Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition* 21(4):803–14. [DAS]
Roediger, H. L., Weldon, M. S. & Challis, B. H. (1989) Explaining dissociation between implicit and explicit measures of retention: A processing account. In: *Varieties of memory and consciousness: Essays in honour of Endel Tulving,* ed. H. L. Roediger & F. I. M. Craik. Erlbaum. [BCL]
Rogers, S. (1995) Perceiving pictorial space. In: *Perception of space and motion,* ed. W. Epstein & S. Rogers. Academic Press. [aHH]
Roitblat, H. L. (1982) The meaning of representation in animal memory. *Behavioral and Brain Sciences* 5:352–72. [MKn]
Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In: *Cognition and the acquisition of language,* ed. T. E. Moore. Academic Press. [KHP]
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M. & Boyes-Braem, P. (1976) Basic objects in natural categories. *Cognitive Psychology* 8:382–439. [arRNS]
Rosh, E. (1978) Principles of categorization. In: *Cognition and categorization,* ed. E. Rosh. LEA. [JRM]
Ross, B. H. (1989) Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15(3):456–68. [DG]
Rucklidge, W. (1996) *Efficient visual recognition using the Hausdorff distance.* Springer-Verlag. [DV]
Ruderman, D. L. (1994) Statistics of natural images. *Network: Computation in Neural Systems* 5:517–48. [aHB]
(1997) Origins of scaling in natural images. *Vision Research* 37:3385–98. [aHB]
Runeson, S. (1974) Constant velocity – not perceived as such. *Psychological Research* 37:3–23. [DMJ]
(1975) Visual prediction of collision with natural and nonnatural motion functions. *Perception and Psychophysics* 18:261–66. [DMJ]
(1977) On the possibility of "smart" perceptual mechanisms. *Scandinavian Journal of Psychology* 18:172–79. [arHH, JH, DMJ, rRNS]
(1988) The distorted room illusion, equilavent configurations, and the specificity of static optic arrays. *Journal of Experimental Psychology: Human Perception and Performance* 14:295–304. [DMJ]
(1989) A note on the utility of ecologically incomplete invariants. *International Society for Ecological Psychology Newsletter* 4(1):6–9. [DMJ]
(1994) Psychophysics: The failure of an elementaristic dream. *Behavioral and Brain Sciences* 17:761–63. [DMJ]
Runeson, S., Jacobs, D. M., Andersson, I. E. K. & Kreegipuu, K. (2001) Specificity is always contingent on constraints; global versus individual arrays is not the issue (Commentary on Stoffregen & Bardy). *Behavioral and Brain Sciences* 21:240–41. [DMJ]
Runeson, S., Juslin, P. & Olsson, H. (2000) Visual perception of dynamic properties: Cue-heuristics versus direct-perceptual competence. *Psychological Review* 107:525–55. [DMJ]
Rushton, S. K. & Wann, J. P. (1999) Weighted combination of size and disparity: A computational model for timing a ball catch. *Nature Neuroscience* 2:186–90. [FL]
Rusin, D. (2000) *The mathematical atlas.* http://www.math.niu.edu/~rusin/known-math/index/55-XX.html [WCH]
Russell, S. (1988) Analogy by similarity. In: *Analogical reasoning,* ed. D. H. Helman. Kluwer Academic. [arRNS, aJBT]
Ryle, G. (1949) *The concept of mind.* Barnes and Noble. [rRS]
Saksida, L. (1999) Effects of similarity and experience on discrimination learning: A nonassociative connectionist model of perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes* 25:308–23. [KC]
Sälström, P. (1973) Colour and physics: Some remarks concerning the physical aspects of human colour vision. *Institute of Physics Report No. 73–09.* Institute of Physics, University of Stockholm. [aRNS]
Saunders, B. & van Brakel, J. (1997) Are there nontrivial constraints on colour categorization? *Behavioral and Brain Sciences* 20:167–228. [LD]
Savelsbergh, G. J. P., Whiting, H. T., Burden, A. M. & Bartlett, R. M. (1992) The role of predictive visual temporal information in the coordination of muscle activity in catching. *Experimental Brain Research* 89:223–28. [FL]
Saxberg, B. V. H. (1987) Projected free fall trajectories: I. Theory and simulation. *Biological Cybernetics* 56:159–75. [aHH]
Schilpp, P. A., ed. (1949) *Albert Einstein: Philosopher-scientist.* The Library of Living Philosophers. [rRNS]
Schirillo, J. A. (1999) Surround articulation. II. Lightness judgements. *Journal of the Optical Society of America A* 16:804–11. [NB]
Schwartz, D. A. (1998) Indexical constraints on symbolic cognitive functioning. In: *Proceedings of the 20th Annual Conference of the Cognitive Science Society,* ed. M. A. Gernsbacher & S. J. Derry. Erlbaum. [DAS]
Schwartz, D. L. (1999) Physical imagery: Kinematic versus dynamic models. *Cognitive Psychology* 38:433–64. [HK]
Schwartz, D. L. & Black, T. (1999) Inferences through imagined actions: Knowing by simulated doing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 25:116–36. [HK]
Schwartz, R. (1969) On knowing a grammar. In: *Language and philosophy,* ed. S. Hook. New York University Press. [rRS]
(1981) Imagery: There's more to it than meets the eye. In: *Imagery,* ed. N. Block. MIT Press. [rRS]
(1984) The problems of representation. *Social Research* 51:1047–64. [rRS]
(1994a) *Vision: Variations on some Berkelean themes.* Blackwell. [rRS]


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

(1994b) Representation. In: *The companion to the philosophy of mind,* ed. S. Guttenplan. Blackwell. [rRS]
(1995) Is mathematical competence innate? *Philosophy of Science* 62:227–40. [rRS]
(1996a) Directed perception. *Philosophical Psychology* 9:81–91. [rRS]
(1996b) Symbols and thought. *Synthese* 106:399–407. [rRS]
(in press) Avoiding error about error. In: *Color perception: From light to object,* ed. R. Mausfeld & D. Heyer. Oxford University Press. [rRS]
Schyns, P., Goldstone, R. L. & Thilbaut, J.-P. (1998) The development of features in subject concepts. *Behavioral and Brain Sciences* 21:1–54. [aJBT]
Scott, D. & Suppes, P. (1958) Foundational aspects of theories of measurement. *Journal of Symbolic Logic* 23:113–28. [aMK]
Selfridge, O. G. (1959) Pandemonium: A paradigm for learning. In: O. G. Selfridge, *Mechanization of thought processes.* Her Majesty's Stationery Office. [DWM]
Shanks, D. R. & Gluck, M. A. (1994) Tests of an adaptive network model for the identification and categorization of continuous-dimension stimuli. *Connection Science* 6:59–89. [aJBT]
Shannon, C. E. (1948) A mathematical theory of communication. *Bell System Technical Journal* 27:379–83, 623–55. [Reprinted in: Shannon, C. E. & Weaver, W. (1949) *The mathematical theory of communication.* University of Illinois Press.] [aRNS]
Shannon, C. E. & Weaver, W., ed. (1949) *The mathematical theory of communication.* University of Illinois Press. [aHB, AR]
Shanon, B. (1976) Ecological constraints in internal representation. *Psychological Review* 91:417–47. [aHH]
Shepard, R. N. (1957) Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika* 22:325–45. [DWM]
(1958) Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review* 65:242–56. [KC, rRNS]
(1962a) The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* 27:219–46. [arRNS]
(1962b) The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika* 27:219–46. [arRNS]
(1964a) Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology* 1:54–87. [aRNS]
(1964b) Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America* 6:2346–53. [aRNS]
(1964c) On subjectively optimum selection among multiattribute alternatives. In: *Human judgments and optimality,* ed. M. Shelley & G. L. Bryan. Wiley. [rRNS]
(1965) Approximation to uniform gradients of generalization by monotone transformations of scale. In: *Stimulus generalization,* ed. D. I. Mostofsky. Stanford University Press. [aRNS]
(1966) Metric structures in ordinal data. *Journal of Mathematical Psychology* 3:287–315. [rRNS]
(1978a) Externalization of mental images and the act of creation. In: *Visual learning, thinking, and communication,* ed. B. S. Randhawa & W. E. Coffman. Academic Press. [rRNS]
(1978b) The circumplex and related topological manifolds in the study of perception. In: *Theory construction and data analysis in the behavioral sciences,* ed. S. Shye. Jossey-Bass. [aRNS]
(1980) Multidimensional scaling, tree-fitting, and clustering. *Science* 210:390–98. [MDL, arRNS, arJBT]
(1981a) Discrimination and classification: A search for psychological laws. Presidential address to the Division of Experimental Psychology of the American Psychological Association, Los Angeles, CA, August 25, 1981. [aRNS]
(1981b) Psychophysical complementarity. In: *Perceptual organization,* ed. M. Kubovy & J. R. Pomerantz. Erlbaum. [WG, aMK, DAS, arRNS, ACZ]
(1981c) Psychophysical relations and psychophysical scales: On the status of "direct" psychophysical measurement. *Journal of Mathematical Psychology* 24:21–57. [rRNS]
(1982a) Geometrical approximations to the structure of musical pitch. *Psychological Review* 89:305–33. [aRNS]
(1982b) Perceptual and analogical bases of cognition. In: *Perspectives on mental representations,* ed. J. Mehler, M. Garrett & E. Walker. Erlbaum. [rRNS]
(1983) Demonstrations of circular components of pitch. *Journal of the Audio Engineering Society* 31:641–49. [aRNS]
(1984) Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review* 91:417–47. [aHB, aHH, HK, aMK, RM, LMP, aRS, arRNS]
(1986) Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General* 115:58–61. [rRNS]
(1987a) Evolution of a mesh between principles of the mind and regularities of the world. In: *The latest on the best: Essays on evolution and optimality,* ed. J. Dupré. MIT Press/Bradford Books. [arMK, RM, JP, arRNS]
(1987b) Toward a universal law of generalization for psychological science. *Science* 237:1317–23. [KC, DD, SE, aHH, EH, MDL, JRM, arRNS, arJBT]
(1988) The role of transformation in spatial cognition. In: *Spatial cognition: Brain bases and development,* ed. J. Stiles-Davis, M. Kritchevsdy & U. Bellugi. Erlbaum. [aRNS]
(1989) A law of generalization and connectionist learning. Plenary address to the Cognitive Science Society, Ann Arbor, MI, August 18, 1989. [aRNS, aJBT]
(1990a) A possible evolutionary basis for trichromacy. In: *Proceedings of the SPIE/SPSE symposium on electronic imaging: Science and technology, vol. 1250. Perceiving, measuring, and using color,* 301–309. [aRNS]
(1990b) *Mind sights: Original visual illusions, ambiguities, and other anomalies, with a commentary on the play of mind in perception and art.* W. H. Freeman. [aMK, rRNS]
(1990c) On understanding mental images. In: *Images and understanding. Thoughts about images; ideas about understanding,* ed. H. Barlow, C. Blakemore & M. Weston-Smith. Cambridge University Press. [aMK]
(1991) Integrality versus separability of stimulus dimensions: From an early convergence of evidence to a proposed theoretical basis. In: *Perception of structure: Essays in honor of Wendell R. Garner,* ed. G. R. Lockhead & J. R. Pomerantz. American Psychological Association. [arRNS]
(1992) The perceptual organization of colors: An adaptation to the regularities of the terrestrial world? In: *The adapted mind: Evolutionary psychology and the generation of culture,* ed. J. Barkow, L. Cosmides & J. Tooby. Oxford University Press. [aRNS, aDT]
(1993) On the physical basis, linguistic representation, and conscious experience of colors. In: *Conceptions of the mind: Essays in honor of George A. Miller,* ed. G. Harman. Erlbaum. [arRNS]
(1994) Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin and Review* 1:2–28. [MB, aHB, NB, DD, aHH, WCH, HI, arMK, GO, aRS, rRNS, aJBT, aDT]
(1995a) Mental universals: Toward a twenty-first century science of mind. In: *The science of the mind: 2001 and beyond,* ed. R. L. Solso & D. W. Massaro. Oxford University Press. [aJBT]
(1995b) What is an agent that it experiences P-consciousness? And what is P-consciousness that it moves an agent? (Commentary on Ned Block). *Behavioral and Brain Sciences* 18:227–87. [rRNS]
(1997) Justification of induction via inverse probabilities and complexity theory. (Unpublished manuscript, Stanford University, January 29, 1997). [rRNS]
(1999) Stream segregation and ambiguity in audition. In: *Music, cognition, and computerized sound: An introduction to psychoacoustics,* ed. P. R. Cook. MIT Press. [rRNS]
Shepard, R. N. & Arabie, P. (1979) Additive clustering representation of similarities as a combination of discrete overlapping properties. *Psychological Review* 86:87–123. [arJBT]
Shepard, R. N. & Carroll, J. D. (1966) Parametric representation of nonlinear data structures. In: *Multivariate analysis,* ed. P. R. Krishnaiah. Academic Press. [aRNS]
Shepard, R. N. & Cermak, G. W. (1973) Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology* 4:351–77. [SE]
Shepard, R. N. & Chang, J. J. (1963) Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology* 65:94–102. [aRNS]
Shepard, R. N. & Chipman, S. (1970) Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology* 1:1–17. [SE, rRNS]
Shepard, R. N. & Cooper, L. A. (1982) *Mental images and their transformations.* MIT Press/Bradford Books. [aMK, arRNS]
(1992) Representation of colors in the blind, color blind, and normally sighted. *Psychological Science* 3:97–104. [aRNS]
Shepard, R. N. & Farrell, J. E. (1985) Representation of the orientations of shapes. *Acta Psychologica* 59:104–21. [arRNS]
Shepard, R. N., Hovland, C. I. & Jenkins, H. M. (1961) Learning and memorization of classifications. *Psychological Monographs* 75 (13, Whole No. 517). [BCL, arRNS]
Shepard, R. N. & Hurwitz, S. (1984) Upward direction, mental rotation, and discrimination of left and right turns in maps. *Cognition* 18:161–93. [rRNS]
Shepard, R. N. & Hut, P. (1998) My experience, your experience, and the world we experience: Turning the hard problem upside down. In: *Toward a science of consciousness II: The Second Tucson Discussions and Debates,* ed. S. R. Hameroff, A. W. Kaszniak & A. C. Scott. MIT Press. [rRNS]
Shepard, R. N. & Judd, S. A. (1976) Perceptual illusion of rotation of three-dimensional objects. *Science* 191:952–54. [TDF, aRNS]
Shepard, R. N. & Kannappan, S. (1991) Toward a connectionist implementation of a theory of generalization. In: *Advances in neural information processing systems, vol. 3,* ed. R. P. Lippmann, J. E. Moody, D. S. Touretzky & S. J. Hanson. Morgan Kaufmann. [arRNS]
Shepard, R. N., Kilpatrick, D. W. & Cunningham, J. P. (1975) The internal representation of numbers. *Cognitive Psychology* 7:82–138. [arRNS, aJBT]



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Shepard, R. N. & Metzler, J. (1971) Mental rotation of three-dimensional objects. *Science* 171:701–703. [TDF, arRNS]
Shepard, R. N. & Tenenbaum, J. (1991) Connectionist modeling of multidimensional generalization. Paper presented at the 32nd annual meeting of the Psychonomic Society, San Francisco, November 22–24, 1991. [arRNS, aJBT]
Shepard, R. N. & Zare, S. (1983) Path-guided apparent motion. *Science* 220:632–34. [arRNS]
Shepard, S. & Metzler, D. (1988) Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance* 14:3–11. [aRNS]
Shiffrar, M. M. & Freyd, J. J. (1990) Apparent motion of the human body. *Psychological Science* 1:257–64 [rRNS]
Shiffrar, M. & Shepard, R. N. (1991) Comparison of cube rotations around axes inclined relative to the environment or to the cube. *Journal of Experimental Psychology: Human Perception and Performance* 17:44–54. [aRS, arRNS]
Shipley, T. F. & Kellman, P. J. (1992) Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception and Psychophysics* 52:97–106. [GV]
Simon, H. A. (1969) *The sciences of the artificial,* (2nd edition 1981). MIT Press. [aHH]
(1990) Invariants of human behavior. *Annual Review of Psychology* 41:1–19. [arHH, rRNS, PMT]
Simons, D. J. & Levin, D. T. (1997) Change blindness. *Trends in Cognitive Science* 1:261–67. [HI]
Singh, M. & Hoffman, D. D. (1999) Completing visual contours: The relationship between relatability and minimizing inflections. *Perception and Psychophysics* 61:943–51. [WG]
Singh, M., Hoffman, D. & Albert, M. (1999) Contour completion and relative depth: Petter's rule and support ratio. *Psychological Science* 10:423–28. [GV]
Sinha, P. & Adelson, E. (1993) Recovering reflectance and illumination in a world of painted polyhedra. *Proceedings of the Fourth International Conference on Computer Vision* (ICCV '93), May 11–14, 1993. [aRNS]
Sinha, P. & Poggio, T. (1996) Role of learning in three-dimensional form perception. *Nature* 384:460–63. [aHB]
Sivik, L. (1997) Color systems for cognitive research. In: *Color categories in thought and language,* ed. C. L. Hardin & L. Maffi. Cambridge University Press. [LD]
Smith, E. E. (1989) Concepts and induction. In: *Foundations of cognitive science,* ed. M. I. Posner. MIT Press. [aJBT]
Smolensky, P. (1988) On the proper treatment of connectionism. *Behavioral and Brain Sciences* 11:1–23. [MDL]
Sokolov, E. N. (1997) Four-dimensional color space. *Behavioral and Brain Sciences* 20:207–208. [LD]
(1998) Higher mental activity and basic physiology: Subjective difference and reaction time. In: *Human cognitive abilities in theory and practice,* ed. J. J. McArdle & R. W. Woodcock. Erlbaum. [ENS]
(2000) Perception and conditioning reflex: Vector encoding. *International Journal of Psychophysiology* 35:197–217. [ENS]
Solomonoff, R. J. (1964a) A formal theory of inductive inference. Part I. *Information and Control* 7:1–22. [aHB, DD]
(1964b) A formal theory of inductive inference. Part II. *Information and Control* 7:24–54. [aHB, DD]
Spelke, E. S. (1991) Physical knowledge in infancy. In: *The epigenesis of mind: Essays on biology and cognition,* ed. S. Carey & R. Gelman. Erlbaum. [arRNS]
(1994) Initial knowledge: Six suggestions. *Cognition* 50:431–45. [HK]
Spelke, E. S., Breinlinger, K., Macomber, J. & Jacobson, K. (1992) Origins of knowledge. *Psychological Review* 99:605–32. [rHH, BH, HK]
Spelke, E. S., Katz, G., Purcell, S. E., Ehrlich, S. M. & Breinlinger, K. (1994) Early knowledge of object motion: Continuity and inertia. *Cognition* 51:131–76. [HK]
Spelke, E. S., Kestenbaum, R., Simons, D. J. & Wein, D. (1995) Spatiotemporal continuity, smoothness of motion, and object identity in infancy. *British Journal of Developmental Psychology* 13:1–30. [HK]
Squire, L. R. (1992) Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review* 99:195–231. [BCL, rJBT]
Stewart, N. & Chater, N. (submitted) The effect of category variability in perceptual categorization. [NC, rJBT]
Sticht, T. G. & Gibson, R. H. (1967) Touch thresholds as a function of onset and offset stimulation. *Psychonomic Science* 8(6):255–56. [DAS]
Streitfeld, B. & Wilson, M. (1986) The ABC's of categorical perception. *Cognitive Psychology* 8:432–51. [KHP]
Sugden, E. H., ed. (1921) *Wesley's standard sermons,* vol. 1. Epworth Press. [rRNS]
Suppes, P. & Zinnes, J. (1963) Basic measurement theory. In: *Handbook of mathematical psychology, vol. I,* ed. R. D. Luce, R. R. Bush & E. Galanter. Wiley. [aMK]
Synge, J. L. (1951) *Science: Sense and nonsense.* Cape. [WCH]
Tenenbaum, J. B. (1996) Learning the structure of similarity. In: *Advances in neural information processing systems, vol. 8,* ed. D. S. Touretzky, M. C. Mozer & M. E. Hasselmo. MIT Press. [aJBT]
(1997) A Bayesian framework for concept learning. In: *SimCat97:Proceedings of the Interdisciplinary Workshop on Similarity and Categorisation* (pp. 249–55), ed. M. Ramscar, U. Hahn, E. Cambouropolos & H. Pain. Department of Artificial Intelligence, Edinburgh University. [arJBT]
(1999a) A Bayesian framework for concept learning. Unpublished doctoral dissertation, Massachussets Institute of Technology, Cambridge, MA. [EH, arJBT]
(1999b) Bayesian modeling of human concept learning. In: *Advances in neural information processing systems, vol. 11,* ed. M. S. Kearns, S. A. Solla & D. A. Cohn. MIT Press. [aJBT]
(2000) Rules and similarity in concept learning. In: *Advances in neural information processing systems, vol. 12,* ed. S. A. Solla, T. K. Leen & K. R. Muller. MIT Press. [arJBT]
Tenenbaum, J. B., de Silva, V. & Langford, J. C. (2000) A global geometric framework for nonlinear dimensionality reduction. *Science* 290:2319–23. [rRNS]
Tenenbaum, J. B. & Griffiths, T. L. (2001) Structure learning in human causal induction. In: *Neural information processing systems 13,* ed. T. K. Leen, T. G. Dietterich & V. Tresp. MIT Press. [rJBT]
Tenenbaum, J. B. & Xu, F. (2000) Word learning as Bayesian inference. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society,* ed. L. R. Gleitman & A. K. Joshi. Erlbaum. [arJBT]
Thompson, E. (1995) *Colour vision.* Routledge. [LD]
Thornber, K. K. & Williams, L. R. (1996) Analytic solution of stochastic completion fields. *Biological Cybernetics* 75:141–51. [GV]
Thurston, W. P. & Weeks, J. R. (1984) The mathematics of three-dimensional manifolds. *Scientific American* 251:108–20. [WCH]
Tinbergen, N. (1963) On aims and methods in ethology. *Zeitschrift für Tierpsychologie* 20:410–33. [KC]
Todd, P. M. (2001) Fast and frugal heuristics for environmentally bounded minds. In: *Bounded rationality: The adaptive toolbox,* ed. G. Gigerenzer & R. Selten. MIT Press. [PMT]
Todd, P. M. & Miller, G. F. (1991) On the sympatric origin of species: Mercurial mating in the quicksilver model. In: *Proceedings of the Fourth International Conference on Genetic Algorithms* (pp. 547–54), ed. R. K. Belew & L. B. Booker. Morgan Kaufmann. [aRNS]
Todorovič, D. (1996) *Is kinematic geometry an internalized regularity?* (Technical Report No. 24/96). Zentrum für interdisziplinare Forschung der Universität Bielefeld. [aMK]
Tolhurst, D. J., Tadmoor, Y. & Chao, T. (1992) Amplitude spectra of natural images. *Ophthalmic and Physiological Optics* 12:229–32. [aHB]
Tolman, E. C. (1948) Cognitive maps in rats and men. *Psychological Review* 55:189–208. [aHB]
Tominaga, S. & Wandell, B. (1989) Standard surface-reflectance model and illuminant estimation. *Journal of the Optical Society of America A* 6:576–84. [aRNS]
Tommasi, L., Bressan, P. & Vallortigara, G. (1995) Solving occlusion indeterminacy in chromatically homogenous patterns. *Perception* 24:391–403. [GV]
Tononi, G., Sporns, O. & Edelman, G. M. (1996) A complexity measure for selective matching of signals by the brain. *Proceedings of the National Academy of Sciences USA* 93:3422–27. [AR]
Tresilian, J. R. (1993) Four questions of time to contact: A critical examination of research on interceptive timing. *Perception* 22:653–80. [FL]
(1999) Visually timed action: Time-out for 'tau'? *Trends in Cognitive Science* 3:301–10. [FL]
Tse, P. U. (1999a) Volume completion. *Cognitive Psychology* 39:37–68. [WG]
(1999b) Complete mergeability and amodal completion. *Acta Psychologica* 102:165–201. [WG]
Tsuda, I. (in press) Towards an interpretation of dynamic neural activity in terms of chaotic dynamical systems. *Behavioral and Brain Sciences.*
Tversky, A. (1977) Features of similarity. *Psychological Review* 84:327–52. [SE, MDL, EMP, rRNS, arJBT]
Tversky, A. & Kahneman, D. (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185:1124–31. [rRNS, rJBT]
(1981) The framing of decisions and the psychology of choice. *Science* 211:453–58. [rRNS]
(1983) Extensional vs. intuitive reasoning: The conjunction fallacy in probabilistic judgment. *Psychological Review* 91:293–315. [rRNS]
Twardy, C. & Bingham, G. P. (1999) The role of energy conservation in event perception. In: *Studies in perception and action V,* ed. M. A. Grealy & J. A. Thomson. Erlbaum. [AW]
(submitted) Causation, causal perception and conservation laws. *Perception and Psychophysics.* [AW]
Ullman, S. (1979) *The interpretation of visual motion.* MIT Press. [aMK, DT]
Van de Vijver, G., Salthe, S. & Delpos, M., eds. (1998) *Evolutionary systems.* Kluwer. [JP]


https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press

Van Gool, L. J., Moons, T., Pauwels, E. & Wagemans, J. (1994) Invariance from the Euclidean geometer's perspective. *Perception* 23:547–61. [DHF]
Van Hateren, F. H. & van der Schaaf, A. (1998) Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society, Series B* 265:359–66. [aHB]
Van Leeuwen, C. & Raffone, A. (2001) Coupled maps as models of perceptual pattern and memory trace dynamics. *Cognitive Processing* 2:67–111. [AR]
Van Leeuwen, C., Steyvers, M. & Nooter, M. (1997) Stability and intermittency in large-scale coupled oscillator models for perceptual segmentation. *Journal of Mathematical Psychology* 41:1–26. [AR]
Vasta, R., Rosenberg, D., Knott, J. A. & Gaze, C. E. (1997) Experience and the water-level-task revisited: Does expertise exact a price? *Psychological Science* 8:336–39. [aHH]
Velichkovsky, B. (1994) The levels approach in psychology and cognitive science. In: *International perspectives on psychological science: Leading themes,* ed. P. Bertelson, P. Eelen & G. d'Ydewalle. Erlbaum. [AH]
Von Foerster, H. (1982) *Observing systems.* Intersystems Publications. [AR]
Von Helmholtz, H. L. F. (1956/1962) Treatise on physiological optics, vol. 3. Dover. [ACZ]
(1866/1965) On empiricism in perception, trans. J. P. C. Southall. In: *A source book in the history of psychology,* ed. R. J. Herrnstein & E. G. Boring. Harvard University Press. (Original work published in 1866). [aMK]
(1894) Über den Ursprung der richtigen Deutung unserer Sinneseindrücke. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane* 7:81–96. [arHH]
(1925) *Physiological optics. Vol. III. The theory of the perception of vision* (Translated from 3rd German edition, 1910). Ch. 26. Optical Society of America. (Dover edition, 1962). [aHB]
Von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q. & Rosander, K. (1998) Principles of predictive action in infancy. *Cognition* 67:255–85. [HK]
Vorobyev, M., Osorio, D., Bennett, A. T., Marshall, N. J. & Cuthill, I. C. (1998) Tetrachromacy, oil droplets and bird plumage colours. *Journal of Comparative Physiology A* 183:621–33. [IG]
Wächtershäuser, G. (1987) Light and life: On the nutritional origins of sensory perception. In: *Evolutionary epistemology, rationality, and the sociology of knowledge,* ed. G. Radnitzky & W. W. Bartely. Open Court. [aHH]
Wallace, C. S. & Boulton, D. M. (1968) An information measure for classification. *Computing Journal* 11:185–95. [aHB, DD]
(1973) An information measure for hierarchic classification. *Computer Journal* 16:254–61. [DD]
Wallace, C. S. & Dowe, D. (1999a) Minimum message length and Kolmogorov complexity. *Computer Journal* 42:270–83. [DD]
(1999b) Refinements of MDL and MML coding. *Computer Journal* 42(4):330–37. [DD]
Wallace, C. S. & Freeman, P. (1987) Estimation and inference by compact coding. *Journal of the Royal Statistical Society B* 49(3):223–65. [DD]
Wallenstein, G. V., Kelso, J. A. S. & Bressler, S. L. (1995) Phase transitions in spatiotemporal patterns of brain activity and behavior. *Physica D* 84:626–34. [TDF]
Walls, G. (1942) *The vertebrate eye and its adaptive radiation.* Cranbrook Institute of Science/Cranbrook Press. [aHB, IG]
Wang, Y. & Frost, B. J. (1992) Time to collision is signalled by neurons in the nucleus rotundus of pigeons. *Nature* 356:236–38. [HK]
Wason, P. C. (1959) The processing of positive and negative information. *Quarterly Journal of Experimental Psychology* 11:92–107. [DAS]
(1966) Reasoning. In: *New horizons in psychology,* ed. B. Foss. Penguin. [DAS]
Wason, P. C. & Johnson-Laird, P. N. (1972) *Psychology of reasoning: Structure and content.* Batsford. [rRNS]
Watanabe, S. (1960) Information-theoretical aspects of inductive and deductive inference. *IBM Journal of Research and Development* 4:208–31. [aHB]
(1985) *Pattern recognition: Human and mechanical.* Wiley. [rRNS]
Watson, A. B. & Ahumada, A. J. (1983) *A look at motion in the frequency domain.* (Tech. Memo 84352). NASA. [MB]
Watson, J. S., Banks, M. S., Von Hofsten, C. & Royden, C. S. (1992) Gravity as a monocular cue for perception of absolute distance and/or absolute size. *Perception* 21:69–76. [aHH]
Wattles, J. (1996) *The golden rule.* Oxford University Press. [rRNS]
Webber, C. J. StC. (1991) Competitive learning, natural images, and the self-organisation of cortical cells. *Network: Computation in Neural Systems* 2:169–87. [aHB]
Weinberg, S. (1992) *Dreams of a final theory.* Pantheon Books. [rRNS]
Weiner, J. (1994) *The beak of the finch: A story of evolution in our time.* Vintage. [LTM]
Werkhoven, P., Snippe, P. & Koenderink, J. (1990) Effects of element orientation on apparent motion perception. *Perception and Psychophysics* 47:509–25. [MB]
Werkhoven, P., Snippe, H. & Toet, A. (1992) Visual processing of optic acceleration. *Vision Research* 32:2313–29. [FL]
Wertheimer, M. (1912) Experimentelle Studien über des das Sehen von Bewegung. *Zeitschrift für Psychologie* 61:161–265. [Translated in part in: T. Shipley, ed. (1961) *Classics in psychology.* Philosophical Library.] [aRNS, aDT]
(1945) *Productive thinking.* Harper. [rRNS]
Westland, S., Shaw, J. & Owens, H. (2000) Colour statistics of natural and man-made surfaces. *Sensor Review* 20:50–55. [NB]
Wiener, N. (1948) *Cybernetics, or control and communication in the animal and the machine.* Wiley. [aRNS]
Wilson, M. (1987) Brain mechanisms in categorical perception. In: *Categorical perception,* ed. S. Harnad. Cambridge University Press. [KHP]
(in press) Perception of imitatable stimuli: Consequences of isomorphism between input and output. *Psychological Bulletin.* [MW]
Wilson, M. & DeBauche, B. A. (1981) Inferotemporal cortex and categorical perception of visual stimuli by monkeys. *Neuropsychologia* 19:29–41. [KHP]
Winfree, A. T. (1980) *The geometry of biological time.* Springer Verlag. [aRNS]
Witkin, H. A. & Goodenough, D. R. (1981) *Cognitive styles: Essence and origins.* International Universities Press. [WCH]
Wittgenstein, L. (1967) *Philosophical investifations, 2nd edition,* trans. G. E. M. Anscombe. Basil Blackwell. (Originally published in 1953). [KKN]
Wolpert, D. H. (1995) Off-training set error and a priori distinctions between learning algorithms. *Santa Fe Institute Technical Report* 95–01–003. [rRNS]
Wolpert, D. H., ed. (1994) *The mathematics of generalization.* Addison Wesley. [rRNS]
Woodworth, R. S. & Sells, S. B. (1935) An atmosphere effect in formal syllogistic reasoning. *Journal of Experimental Psychology* 18:451–60. [rRNS]
Wunderlich, H. (1977) *Kursächsische Feldmeßkunst, artilleristische Richtverfahren und Ballistik im 16. und 17. Jahrhundert.* VEB Deutscher Verlag der Wissenschaften. [aHH]
Wyszecki, G. & Stiles, W. S. (1982) *Color science, second edition.* Wiley. [MHB, LD]
Yamauchi, T. & Markman, A. B. (1998) Category learning by inference and classification. *Journal of Memory and Language* 37:124–49. [BCL]
Young, M. P. & Yamane, S. (1992) Sparse population coding of faces in the inferotemporal cortex. *Science* 256:1327–30. [AH]
Zadeh, L. A. (1982) A note on prototype theory and fuzzy sets. *Cognition* 12:291–97. [DWM]
Zahavi, A. & Zahavi, A. (1997) *The handicap principle: A missing piece of Darwin's puzzle.* Oxford University Press. [PMT]
Zanforlin, M. (1976) Observations on the visual perception of the snail *Euparipha pisana* (Muller). *Bolletino di Zoologia* 43:303–15. [GV]
Zera, J. & Green, D. M. (1993) Detecting temporal onset and offset asynchrony in multicomponent complexes. *Journal of the Acoustical Society of America* 93(2):1038–52. [DAS]
Zipser, K., Lamme, A. F. & Schiller, P. H. (1996) Contextual modulation in primary visual cortex. *Journal of Neuroscience* 16:7376–89. [AR]



https://doi.org/10.1017/S0140525X01620089 Published online by Cambridge University Press