BEHAVIORAL AND BRAIN SCIENCES (2007) **30**, 1–61
Printed in the United States of America
DOI: 10.1017/S0140525X07000581

A framework for the unification of the behavioral sciences

Herbert Gintis
Behavioral Sciences, Santa Fe Institute, Santa Fe, NM 87501; Department of Economics, Central European University, Budapest, H-1051 Hungary
hgintis@comcast.net **http://www-unix.oit.umass.edu/~gintis**

Abstract: The various behavioral disciplines model human behavior in distinct and incompatible ways. Yet, recent theoretical and empirical developments have created the conditions for rendering coherent the areas of overlap of the various behavioral disciplines. The analytical tools deployed in this task incorporate core principles from several behavioral disciplines. The proposed framework recognizes evolutionary theory, covering both genetic and cultural evolution, as the integrating principle of behavioral science. Moreover, if decision theory and game theory are broadened to encompass other-regarding preferences, they become capable of modeling all aspects of decision making, including those normally considered "psychological," "sociological," or "anthropological." The mind as a decision-making organ then becomes the organizing principle of psychology.

Keywords: behavioral game theory; behavioral science; evolutionary theory; experimental psychology; gene-culture coevolution; rational actor model; socialization

1. Introduction

The behavioral sciences encompass economics, biology, anthropology, sociology, psychology, and political science, as well as their subdisciplines, including neuroscience, archaeology, and paleontology, and to a lesser extent, such related disciplines as history, legal studies, and philosophy.[1] These disciplines have many distinct concerns, but each includes a model of individual human behavior. These models are not only different, which is to be expected given their distinct explanatory goals, but also *incompatible*. Nor can this incompatibility be accounted for by the type of causality involved (e.g., *ultimate* as opposed to *proximate* explanations). This situation is well known but does not appear discomforting to behavioral scientists, as there has been virtually no effort to repair this condition.[2] In their current state, however, according the behavioral sciences the status of true sciences is less than credible.

One of the great triumphs of twentieth-century science was the seamless integration of physics, chemistry, and astronomy, on the basis of a common model of fundamental particles and the structure of space-time. Of course, gravity and the other fundamental forces, which operate on extremely different energy scales, have yet to be reconciled; and physicists are often criticized for their seemingly endless generating of speculative models that might accomplish this reconciliation. But a similar dissatisfaction with analytical incongruence on the part of their practitioners would serve the behavioral sciences well. This paper argues that we now have the analytical and empirical bases to construct the framework for an integrated behavioral science.

The behavioral sciences all include models of individual human behavior. These models should be compatible. Indeed, there should be a common underlying model, enriched in different ways to meet the particular needs of each discipline. We cannot easily attain this goal at present, however, as the various behavioral disciplines currently have *incompatible* models. Yet, recent theoretical and empirical developments have created the conditions for rendering coherent the areas of overlap of the various behavioral disciplines. The analytical tools deployed in this task incorporate core principles from several behavioral disciplines.[3]

The standard justification for the fragmentation of the behavioral disciplines is that each has a model of human behavior well suited to its particular object of study. While this is true, where these objects of study *overlap*, their models must be compatible. In particular, psychology, economics, anthropology, biology, and sociology should have concordant explanations of law-abiding behavior, charitable giving, political corruption, voting behavior, and other complex behaviors that do not fit nicely within disciplinary boundaries. They do not have such explanations presently.

This paper sketches a framework for the unification of the behavioral sciences. Two major conceptual categories – evolution and game theory – cover *ultimate* and *proximate* causality. Under each category are conceptual subcategories that relate to overlapping interests of two or more behavioral disciplines. In this target article, I argue the following points:

1.1. Evolutionary perspective

Evolutionary biology underlies all behavioral disciplines because *Homo sapiens* is an evolved species whose characteristics are the product of its particular evolutionary history.

1.1.1. Gene-culture coevolution. The centrality of culture and complex social organization to the evolutionary success of *Homo sapiens* implies that fitness in humans will depend

© 2007 Cambridge University Press 0140-525X/07 $40.00

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

on the structure of cultural life.[4] Because culture is influenced by human genetic propensities, it follows that human cognitive, affective, and moral capacities are the products of a unique dynamic known as *gene-culture coevolution, in which genes adapt to a fitness landscape of which cultural forms are a critical element, and the resulting genetic changes lay the basis for further cultural evolution*. This coevolutionary process has endowed us with preferences that go beyond the self-regarding concerns emphasized in traditional economic and biological theories, and embrace such other-regarding values as a taste for cooperation, fairness, and retribution; the capacity to empathize; and the ability to value such constitutive behaviors as honesty, hard work, toleration of diversity, and loyalty to one's reference group.[5]

1.1.2. Imitation and conformist transmission. Cultural transmission generally takes the form of *conformism* – that is, individuals accept the dominant cultural forms, ostensibly because it is fitness-enhancing to do so (Bandura 1977; Boyd & Richerson 1985; Conlisk 1988; Krueger & Funder 2004). Although adopting the beliefs, techniques, and cultural practices of successful individuals is a major mechanism of cultural transmission, there is constant cultural mutation, and individuals may adopt new cultural forms when those forms appear better to serve their interests (Gintis 1972; 2003b; Henrich 2001). One might expect that the analytical apparatus for understanding cultural transmission, including the evolution, diffusion, and extinction of cultural forms, might come from sociology or anthropology, the disciplines that focus on cultural life; but such is not the case. Both fields treat culture in a static manner that belies its dynamic and evolutionary character. By recognizing the common nature of genes and culture as forms of information that are transmitted intergenerationally, biology offers an analytical basis for understanding cultural transmission.

HERBERT GINTIS is an External Faculty member at the Santa Fe Institute, New Mexico, and Professor of Economics at the Central European University, Budapest, Hungary. He heads a multidisciplinary research project that models such behaviors as empathy, reciprocity, insider/outsider behavior, vengefulness, and other observed human behaviors not well handled by the traditional model of the self-regarding agent. His web site, www-unix.oit.umass.edu/~gintis, contains pertinent information. He has published *Game Theory Evolving* (Princeton University Press, 2000), and is coeditor of *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-scale Societies* (Oxford University Press, 2004) and *Moral Sentiments and Material Interests: On the Foundations of Cooperation in Economic Life* (MIT Press, 2005). He is currently completing a new book titled *A Cooperative Species: Human Reciprocity and its Evolution*.

1.1.3. Internalization of norms. In sharp contrast to other species, humans have preferences that are *socially programmable*, in the sense that the individual's goals, and not merely the methods for their satisfaction, are acquired through a social learning process. Culture therefore takes the form not only of new techniques for controlling nature, but also of *norms and values* that are incorporated into individual preference functions through the sociological mechanism known as *socialization* and the psychological mechanism known as *internalization of norms*. Surprisingly, the internalization of norms, which is perhaps the most singularly characteristic feature of the human mind and central to understanding cooperation and conflict in human society, is ignored or misrepresented in the other behavioral disciplines, anthropology and social psychology aside.

1.2. Game theory

The analysis of living systems includes one concept that does not occur in the nonliving world and that is not analytically represented in the natural sciences. This is the notion of a *strategic interaction*, in which the behavior of individuals is derived by assuming that each is choosing a *fitness-relevant response* to the actions of other individuals. The study of systems in which individuals choose fitness-relevant responses and in which such responses evolve dynamically, is called *evolutionary game theory*. Game theory provides a transdisciplinary conceptual basis for analyzing choice in the presence of strategic interaction. However, the classical game-theoretic assumption that individuals are self-regarding must be abandoned except in specific situations (e.g., anonymous market interactions), and many characteristics that classical game theorists have considered logical implications of the principles of rational behavior, including the use of backward induction, are in fact not implied by rationality. Reliance on classical game theory has led economists and psychologists to mischaracterize many common human behaviors as irrational. Evolutionary game theory, whose equilibrium concept is that of a stable stationary point of a dynamical system, must therefore replace classical game theory, which erroneously favors subgame perfection and sequentiality as equilibrium concepts.

1.2.1. The brain as a decision-making organ. In any organism with a central nervous system, the brain evolved because centralized information processing entailed enhanced decision-making capacity, the fitness benefits thereof more than offsetting its metabolic and other costs. Therefore, decision making must be the central organizing principle of psychology. This is not to say that learning (the focus of behavioral psychology) and information processing (the focus of cognitive psychology) are not of supreme importance, but rather, that principles of learning and information processing only make sense in the context of the decision-making role of the brain.[6]

1.2.2. The rational actor model. General evolutionary principles suggest that individual decision-making can be modeled as optimizing a preference function subject to informational and material constraints. Natural selection ensures that content of preferences will reflect biological fitness, at least in the environments in which preferences evolved. The principle of expected utility extends this optimization to stochastic outcomes. The resulting model is called the *rational actor model* in economics, but I will generally refer to this as the *beliefs, preferences, and*

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

constraints (BPC) *model* to avoid the often misleading connotations attached to the term "rational."[7]

Economics, biology, and political science integrate game theory into the core of their models of human behavior. By contrast, game theory widely evokes emotions, from laughter to hostility, in other behavioral disciplines. Certainly, if one rejects the BPC model (as these other disciplines characteristically do), game theory makes no sense whatsoever. The standard critiques of game theory in these other disciplines are indeed generally based on the sorts of arguments on which the critique of the BPC model is based; I discuss this in section 9.

In addition to these conceptual tools, the behavioral sciences of course share common access to the natural sciences, statistical and mathematical techniques, computer modeling, and a common scientific method.

The afore-mentioned principles are certainly not exhaustive. The list is quite spare, and will doubtless be expanded in the future. Note that I am not asserting that the above-mentioned principles are *the most important* in each behavioral discipline. Rather, I am saying that they contribute to constructing a bridge across disciplines—a common model of human behavior from which each discipline can branch.

Accepting the above framework may entail substantive reworking of basic theory in a particular discipline, but I expect that much research will be relatively unaffected by this reworking. For instance, a psychologist working on visual processing, or an economist working on futures markets, or an anthropologist tracking food-sharing practices across social groups, or a sociologist gauging the effect of dual parenting on children's educational attainment, might gain little from knowing that a unified model underlay all the behavioral disciplines. But, I suggest that in such critical areas as the relationship between corruption and economic growth, community organization and substance abuse, taxation and public support for the welfare state, and the dynamics of criminality, researchers in one discipline are likely to benefit greatly from interacting with sister disciplines in developing valid and useful models.

In what follows, I expand on each of the above concepts, after which I address common objections to the beliefs, preferences, and constraints (BPC) model and game theory.

2. Evolutionary perspective

A *replicator* is a physical system capable of making copies of itself. Chemical crystals, such as salt, have this property of replication; but biological replicators have the additional ability to assume a myriad of physical forms based on the highly variable sequencing of their chemical building blocks (Schrödinger called life an "aperiodic crystal" in 1944, before the structure of DNA was discovered). Biology studies the dynamics of such complex replicators, using the evolutionary concepts of replication, variation, mutation, and selection (Lewontin 1974).

Biology plays a role in the behavioral sciences much like that of physics in the natural sciences. Just as physics studies the elementary processes that underlie all natural systems, so biology studies the general characteristics of survivors of the process of natural selection. In particular, genetic replicators, the environments to which they give rise, and the effect of these environments on gene frequencies, account for the characteristics of species, including the development of individual traits and the nature of intraspecific interaction. This does not mean, of course, that behavioral science in any sense *reduces* to biological laws. Just as one cannot deduce the character of natural systems (e.g., the principles of inorganic and organic chemistry, the structure and history of the universe, robotics, plate tectonics) from the basic laws of physics, similarly one cannot deduce the structure and dynamics of complex life forms from basic biological principles. But, just as physical principles inform model creation in the natural sciences, so must biological principles inform all the behavioral sciences.

3. The brain as a decision-making organ

The fitness of an organism depends on how effectively it makes choices in an uncertain environment. Effective choice must be a function of the organism's state of knowledge, which consists of the information supplied by the sensory organs that monitor the organism's internal states and its external environment. In relatively simple organisms, the choice environment is primitive and distributed in a decentralized manner over sensory inputs. But, in three separate groups of animals – the craniates (vertebrates and related creatures), arthropods (including insects, spiders, and crustaceans), and cephalopods (squid, octopuses, and other mollusks) – a central nervous system with a brain (a centrally located decision-making and control apparatus) evolved. The phylogenetic tree of vertebrates exhibits increasing complexity through time, and increasing metabolic and morphological costs of maintaining brain activity. *The brain evolved because more complex brains, despite their costs, enhanced the fitness of their bearers*. Brains, therefore, are ineluctably structured to make, on balance, fitness-enhancing decisions in the face of the various constellations of sensory inputs their bearers commonly experience.

The human brain shares most of its functions with that of other vertebrate species, including the coordination of movement, maintenance of homeostatic bodily functions, memory, attention, processing of sensory inputs, and elementary learning mechanisms. The distinguishing characteristic of the human brain, however, lies in its extraordinary power as a *decision-making* mechanism.

Surprisingly, this basic insight is missing from psychology, which focuses on the processes that render decision-making possible (attention, logical inference, emotion vs. reason, categorization, relevance) but virtually ignores, and seriously misrepresents decision making itself. Psychology has two main branches: cognitive and behavioral. The former defines the brain as an "information-processing organ" and generally argues that humans are relatively poor, irrational, and inconsistent decision makers. The latter is preoccupied with learning mechanisms that humans share with virtually all metazoans (stimulus response, the law of effect, operant conditioning, and the like). For example, a widely used text of graduate-level readings in cognitive psychology (Sternberg & Wagner 1999) devotes the *ninth* of 11 chapters to "Reasoning, Judgment, and Decision Making." It offers two papers, the first of which shows that human subjects generally fail simple

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

logical inference tasks, and the second shows that human subjects are irrationally swayed by the way a problem is verbally "framed" by the experimenter. A leading undergraduate cognitive psychology text (Goldstein 2005) places "Reasoning and Decision Making" *last* among 12 chapters. This chapter includes one paragraph describing the rational actor model, followed by many pages purporting to explain why the model is wrong. Behavioral psychology generally avoids positing internal states, of which beliefs and preferences and even some constraints (e.g., a character virtue such as keeping promises) are examples. When the rational actor model is mentioned with regard to human behavior, it is summarily rejected (Herrnstein et al. 1997). Not surprisingly, in a leading behavioral psychology text (Mazur 2002), choice is covered in the *last* of 14 chapters and is limited to a review of the literature on choice between concurrent reinforcement schedules and the capacity to defer gratification.

Summing up a quarter century of psychological research in 1995, Paul Slovic asserted, accurately I believe, that "it is now generally recognized among psychologists that utility maximization provides only limited insight into the processes by which decisions are made" (Slovic 1995, p. 365). "People are not logical," psychologists are fond of saying, "they are *psycho*logical." In this paper I argue precisely the opposite position: people are generally rational, though subject to performance errors.

Psychology could be the centerpiece of the human behavioral sciences by providing a general model of decision making for the other behavioral disciplines to use and elaborate for their various purposes. The field fails to hold this position because its core theories do not take the fitness-enhancing character of the human brain, its capacity to make effective decisions in complex environments, as central.[8]

4. The foundations of the BPC model

For every constellation of sensory inputs, each decision taken by an organism generates a probability distribution over fitness outcomes, the expected value of which is the *fitness* associated with that decision. Because fitness is a scalar variable (basically, the expected number of offspring to reach reproductive maturity), for each constellation of sensory inputs, each possible action the organism might take has a specific fitness value; and organisms whose decision mechanisms are optimized for this environment will choose the available action that maximizes this fitness value.[9] It follows that, given the state of its sensory inputs, if an organism with an optimized brain chooses action A over action B when both are available, and chooses action B over action C when both are available, then it will also choose action A over action C when both are available. This is called *choice consistency*.

The rational actor model was developed in the twentieth century by John von Neumann, Leonard Savage, and many others. The model appears prima facie to apply only when individuals can determine all the logical and mathematical implications of the knowledge they possess. However, the model in fact depends only on choice consistency and the assumption that an individual can trade off among outcomes, in the sense that for any finite set of outcomes $A_1 \ldots A_n$, if A_1 is the least preferred and A_n the most preferred outcome, then for any A_i, where $1 \le i \le n$, there is a probability p_i, where $0 \le p_i \le 1$, such that the individual is indifferent between A_i and a lottery that pays A_1 with probability p_i and pays A_n with probability $1-p_i$ (Kreps 1990). (A *lottery* is a probability distribution over a finite set of outcomes.) Clearly, these assumptions are often extremely plausible. When applicable, the rational actor model's choice consistency assumption enhances explanatory power, even in areas that have traditionally rejected the model (Coleman 1990; Hechter & Kanazawa 1997; Kollock 1997). In short, when preferences are consistent, they can be represented by a numerical function, which we call the objective function, that individuals maximize subject to their beliefs (including Bayesian probabilities) and the constraints that they face.

Four caveats are in order. First, this analysis does not suggest that people consciously maximize anything. Second, the model does *not* assume that individual choices, even if they are self-referring (e.g., personal consumption), are always welfare-enhancing. Third, preferences must be stable across time to be theoretically useful; but preferences are ineluctably a function of such parameters as hunger, fear, and recent social experience, and beliefs can change dramatically in response to immediate sensory experience. Finally, the BPC model does not presume that beliefs are correct or that they are updated correctly in the face of new evidence, although Bayesian assumptions concerning updating can be made part of preference consistency in elegant and compelling ways (Jaynes 2003).

The rational actor model is the cornerstone of contemporary economic theory, and in the past few decades it has become equally important in the biological modeling of animal behavior (Alcock 1993; Real 1991; Real & Caraco 1986). Economic and biological theory therefore have a natural affinity: The choice consistency on which the rational actor model of economic theory depends is rendered plausible by biological evolutionary theory, and the optimization techniques pioneered by economic theorists are routinely applied and extended by biologists in modeling the behavior of organisms.

For similar reasons, in a stochastic environment, natural selection will enhance the capacity of the brain to make choices that enhance expected fitness, and hence that satisfy the expected utility principle. To see this, suppose an organism must choose from action set X, where each $x \in X$ determines a lottery that pays i offspring with probability $p_i(x)$, for $i = 0, 1, \ldots, n$. Then the expected number of offspring from this lottery is

$$\psi(x) = \sum_{j=1}^{n} j p_j(x)$$

Let L be a lottery on X that delivers $x_i \in X$ with probability q_i for $i = 1, \ldots, k$. The probability of j offspring given L is then

$$\sum_{i=1}^{k} q_i p_i(x_i)$$

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

so the expected number of offspring given L is

$$\sum_{j=1}^{n}\sum_{i=1}^{k} q_i p_j(x_i) = \sum_{i=1}^{k} q_i \sum_{i=1}^{k} j p_j(x_i)$$

$$= \sum_{i=1}^{k} q_i \psi(x_i)$$

which is the expected value theorem with utility function $\psi(\bullet)$. (See also Cooper 1987.)

There are few reported failures of the expected utility theorem in nonhumans, and there are some compelling examples of its satisfaction (Real & Caraco 1986). The difference between humans and other animals is that the latter are tested in *real life*, or in elaborate simulations of real life, whereas humans are tested in the laboratory under conditions differing radically from real life. Although it is important to know how humans choose in such situations (see sect. 9), there is certainly no guarantee they will make the same choices in the real-life situation and in the situation analytically generated to represent it. For example, a heuristic that says "adopt choice behavior that appears to have benefited others" may lead to expected maximization even when individuals are error-prone when evaluating stochastic alternatives in the laboratory.

In addition to the explanatory success of theories based on the rational actor model, supporting evidence from contemporary neuroscience suggests that maximization is not simply an "as if" story. In fact, the brain's neural circuitry makes choices by internally representing the payoffs of various alternatives as neural firing rates, choosing such a maximal rate (Dorris & Glimcher 2004; Glimcher 2003; Glimcher et al. 2005). Neuroscientists increasingly find that an aggregate decision-making process in the brain synthesizes all available information into a single, unitary value (Glimcher 2003; Parker & Newsome 1998; Schall & Thompson 1999). Indeed, when animals are tested in a repeated trial setting with variable reward, dopamine neurons appear to encode the difference between the reward that an animal expected to receive and the reward that an animal actually received on a particular trial (Schultz et al. 1997; Sutton & Barto 2000), an evaluation mechanism that enhances the environmental sensitivity of the animal's decision-making system. This error-prediction mechanism has the drawback of seeking only local optima (Sugrue et al. 2005). Montague and Berns (2002) address this problem, showing that the orbitofrontal cortex and striatum contain mechanisms for more global predictions that include risk assessment and discounting of future rewards. Their data suggest a decision-making model that is analogous to the famous Black-Scholes options pricing equation (Black & Scholes 1973).

Although the neuroscientific evidence supports the BPC model, it does not support the traditional economic model of Homo economicus. For instance, recent evidence supplies a neurological basis for hyperbolic discounting, and hence undermines the traditional belief in time consistent preferences. McClure et al. (2004) showed that two separate systems are involved in long- versus short-term decisions. The lateral prefrontal cortex and posterior parietal cortex are engaged in all intertemporal choices, while the paralimbic cortex and related parts of the limbic system kick in only when immediate rewards are available. Indeed, the relative engagement of the two systems is directly associated with the subject's relative favoring of long-term over short-term reward.

The BPC model is the most powerful analytical tool of the behavioral sciences. For most of its existence this model has been justified in terms of "revealed preferences," rather than by the identification of neural processes that generate constrained optimal outcomes. The neuroscience evidence suggests a firmer foundation for the rational actor model.

5. Gene-culture coevolution

The genome encodes information that is used to construct a new organism, to instruct the new organism how to transform sensory inputs into decision outputs (i.e., to endow the new organism with a specific preference structure), and to transmit this coded information virtually intact to the new organism. Because learning about one's environment may be costly and is error prone, efficient information transmission will ensure that the genome encodes all aspects of the organism's environment that are constant, or that change only very slowly through time and space. By contrast, environmental conditions that vary across generations and/or in the course of the organism's life history can be dealt with by providing the organism with the capacity to *learn*, and hence phenotypically adapt to specific environmental conditions.

There is an intermediate case that is not efficiently handled by either genetic encoding or learning. When environmental conditions are positively but imperfectly correlated across generations, each generation acquires valuable information through learning that it cannot transmit genetically to the succeeding generation, because such information is not encoded in the germ line. In the context of such environments, there is a fitness benefit to the transmission of information by means other than the germ line concerning the current state of the environment. Such *epigenetic* information is quite common (Jablonka & Lamb 1995), but it achieves its highest and most flexible form in *cultural transmission* in humans and, to a lesser extent, in primates and other animals (Bonner 1984; Richerson & Boyd 1998). Cultural transmission takes the form of vertical (parents to children), horizontal (peer to peer), and oblique (elder to younger), as in Cavalli-Sforza and Feldman (1981); prestige (higher-influencing lower-status), as in Henrich and Gil-White (2001); popularity-related, as in Newman et al. (2006); and even random population-dynamic transmission, as in Shennan (1997) and Skibo and Bentley (2003).

The parallel between cultural and biological evolution goes back to Huxley (1955), Popper (1979), and James (1880).[10] The idea of treating culture as a form of epigenetic transmission was pioneered by Richard Dawkins, who coined the term "meme" in *The Selfish Gene* (1976) to represent an integral unit of information that could be transmitted phenotypically. There quickly followed several major contributions to a biological approach to culture, all based on the notion that culture, like genes, could evolve through replication (intergenerational transmission), mutation, and selection (Boyd & Richerson 1985; Cavalli-Sforza & Feldman 1982; Lumsden & Wilson 1981).

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Cultural elements reproduce themselves from brain to brain and across time, mutate, and are subject to selection according to their effects on the fitness of their carriers (Boyd & Richerson 1985; Cavalli-Sforza & Feldman 1982; Parsons 1964). Moreover, there are strong interactions between genetic and epigenetic elements in human evolution, ranging from basic physiology (e.g., the transformation of the organs of speech with the evolution of language) to sophisticated social emotions, including empathy, shame, guilt, and revenge seeking (Zajonc 1980; 1984).

Because of their common informational and evolutionary character, there are strong parallels between genetic and cultural modeling (Mesoudi et al. 2006). Like biological transmission, culture is transmitted from parents to offspring; and like cultural transmission, wherein culture is transmitted *horizontally* among *unrelated* individuals, so too in microbes and many plant species genes are regularly transferred across lineage boundaries (Abbott et al. 2003; Jablonka & Lamb 1995; Rivera & Lake 2004). Moreover, anthropologists reconstruct the history of social groups by analyzing homologous and analogous cultural traits, much as biologists reconstruct the evolution of species by the analysis of shared characters and homologous DNA (Mace & Pagel 1994). Indeed, the same computer programs developed by biological systematists are used by cultural anthropologists (Holden 2002; Holden & Mace 2003). In addition, archeologists who study cultural evolution have a modus operandi similar to that of paleobiologists who study genetic evolution (Mesoudi et al. 2006): both attempt to reconstruct lineages of artifacts and their carriers. Like paleobiology, archaeology assumes that when analogy can be ruled out, similarity implies causal connection by inheritance (O'Brien & Lyman 2000). Like biogeography's study of the spatial distribution of organisms (Brown & Lomolino 1998), behavioral ecology studies the interaction of ecological, historical, and geographical factors that determine distribution of cultural forms across space and time (Smith & Winterhalder 1992).

Perhaps the most common critique of the analogy between genetic and cultural evolution is that the gene is a well-defined, distinct, independently reproducing and mutating entity, whereas the boundaries of the unit of culture are ill-defined and overlapping. In fact, however, this view of the gene is simply outdated. Overlapping, nested, and movable genes, discovered in the course of the past 35 years, have some of the fluidity of cultural units, whereas often the boundaries of a cultural unit (a belief, icon, word, technique, stylistic convention) are quite delimited and specific. Similarly, alternative splicing, nuclear and messenger RNA editing, cellular protein modification, and genomic imprinting – which are quite common – undermine the standard view of the insular gene producing a single protein, and support the notion of genes having variable boundaries and strongly context-dependent effects.

Dawkins added a second fundamental mechanism of epigenetic information transmission in *The Extended Phenotype* (Dawkins 1982), noting that organisms can directly transmit environmental artifacts to the next generation, in the form of such constructs as beaver dams, beehives, and even social structures (e.g., mating and hunting practices). The phenomenon of a species creating an important aspect of its environment and stably transmitting this environment across generations, known as *niche construction*, is a widespread form of epigenetic transmission (Odling-Smee et al. 2003). Moreover, niche construction gives rise to what might be called a *gene–environment coevolutionary process* – that is, a genetically induced environmental regularity becomes the basis for genetic selection, and genetic mutations that give rise to mutant niches survive if they are fitness-enhancing for their constructors. The dynamical modeling of the reciprocal action of genes and culture is known as *gene–culture coevolution* (Bowles & Gintis 2005a; Durham 1991; Feldman & Zhivotovsky 1992; Lumsden & Wilson 1981).

An excellent example of gene–environment coevolution is the honeybee, in which the origin of its eusociality doubtless lay in the high degree of relatedness fostered by haplodiploidy, but which persists in modern species despite the fact that relatedness in the hive is generally quite low, on account of multiple queen matings, multiple queens, queen deaths, and the like (Gadagkar 1991; Seeley 1997). The social structure of the hive is transmitted epigenetically across generations, and the honeybee genome is an adaptation to the social structure laid down in the distant past.

Gene–culture coevolution in humans is a special case of gene–environment coevolution in which the environment is culturally constituted and transmitted (Feldman & Zhivotovsky 1992). The key to the success of our species in the framework of the hunter-gatherer social structure in which we evolved is the capacity of unrelated, or only loosely related, individuals to cooperate in relatively large egalitarian groups in hunting and territorial acquisition and defense (Boehm 2000; Richerson & Boyd 2004). Although contemporary biological and economic theory have attempted to show that such cooperation can be effected by self-regarding rational agents (Alexander 1987; Fudenberg et al. 1994; Trivers 1971), the conditions under which this is the case are highly implausible even for small groups (Boyd & Richerson 1988; Gintis 2005b). Rather, the social environment of early humans was conducive to the development of prosocial traits, such as empathy, shame, pride, embarrassment, and reciprocity, without which social cooperation would be impossible.

Neuroscientific studies exhibit clearly both the neural plasticity of and the genetic basis for moral behavior. Brain regions involved in moral judgments and behavior include the prefrontal cortex, the orbitofrontal cortex, and the superior temporal sulcus (Moll et al. 2005). These brain structures are present in all primates but are most highly developed in humans and are doubtless evolutionary adaptations (Schulkin 2000). The evolution of the human prefrontal cortex is closely tied to the emergence of human morality (Allman et al. 2002). Patients with focal damage to one or more of these areas exhibit a variety of antisocial behaviors, including sociopathy (Miller et al. 1997) and the absence of embarrassment, pride, and regret (Beer et al. 2003; Camille 2004).

6. The concept of culture across disciplines

Because of the centrality of culture to the behavioral sciences, it is worth noting the divergent use of the

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

concept in distinct disciplines, and the sense in which it is used here.

Anthropology, the discipline that is most sensitive to the vast array of cultural groupings in human societies, treats culture as an expressive totality defining the life space of individuals, including symbols, language, beliefs, rituals, and values.

By contrast, in biology, culture is generally treated as *information*, in the form of instrumental techniques and practices, such as those used in producing necessities, fabricating tools, waging war, defending territory, maintaining health, and rearing children. We may include in this category *conventions* (e.g., standard greetings, forms of dress, rules governing the division of labor, the regulation of marriage, and rituals) that differ across groups and serve to coordinate group behavior, facilitate communication and the maintenance of shared understandings. Similarly, we may include *transcendental beliefs* (e.g., that sickness is caused by angering the gods, that good deeds are rewarded in the afterlife) as a form of information. A transcendental belief is the assertion of a state of affairs that has a truth value, but one that believers either cannot or choose not to test personally (Atran 2004). Cultural transmission in humans, in this view, is thus a process of information transmission, rendered possible by our uniquely prodigious cognitive capacities (Tomasello et al. 2005).

The predisposition of a new member to accept the dominant cultural forms of a group is called *conformist transmission* (Boyd & Richerson 1985). Conformist transmission may be fitness-enhancing, because, if an individual must determine the most effective of several alternative techniques or practices, and if experimentation is costly, it may be payoff-maximizing to copy others rather than incur the costs of experimenting (Boyd & Richerson 1985; Conlisk 1988). Conformist transmission extends to the transmission of transcendental beliefs, as well. Such beliefs affirm techniques where the cost of experimentation is extremely high or infinite, and the cost of making errors is also high. This is, in effect, Blaise Pascal's argument for believing in God. This view of religion is supported by Boyer (2001), who models transcendental beliefs as cognitive beliefs that coexist and interact with our other more mundane beliefs. In this view, one conforms to transcendental beliefs because their truth value has been ascertained by others (relatives, ancestors, prophets) and are deemed to be as worthy of affirmation as the everyday techniques and practices, such as norms of personal hygiene, that one accepts on faith, without personal verification.

Sociology and anthropology recognize the importance of conformist transmission but the notion is virtually absent from economic theory. For instance, in economic theory consumers maximize utility and firms maximize profits by considering only market prices and their own preference and production functions. In fact, in the face of incomplete information and the high cost of information-gathering, both consumers and firms in the first instance may simply imitate what appear to be the successful practices of others, adjust their behavior incrementally in the face of varying market conditions, and sporadically inspect alternative strategies in limited areas (Gintis in press a; 2006c).

Possibly part of the reason the BPC model is so widely rejected in some disciplines is because of the belief that optimization is analytically incompatible with reliance on imitation and hence with conformist transmission. In fact, the economists' distaste for optimization *via* imitation is not complete (Bikhchandani et al. 1992; Conlisk 1988). Recognizing that imitation is an aspect of optimization has the added attractiveness of allowing us to model cultural change in a dynamic manner: New cultural forms displace older forms when they appear to advance the goals of their bearers (Gintis 2003b; Henrich 1997; 2001; Henrich & Boyd 1998).

7. Programmable preferences and the sociology of choice

Sociology, in contrast to biology, treats culture primarily as a set of *moral values* (e.g., norms of fairness, reciprocity, justice) that are held in common by members of the community (or a stratum within the community) and are transmitted from generation to generation by the process of *socialization*. According to Durkheim (1951), the organization of society involves assigning individuals to specific *roles*, each with its own set of socially sanctioned values. A key tenet of socialization theory is that a society's values are passed from generation to generation through the *internalization of norms* (Benedict 1934; Durkheim 1951; Grusec & Kuczynski 1997; Mead 1963; Nisbett & Cohen 1996; Parsons 1967; Rozin et al. 1999), which is a process in which the initiated instill values into the uninitiated (usually the younger generation) through an extended series of personal interactions, relying on a complex interplay of affect and authority. Through the internalization of norms, initiates are supplied with moral values that induce them to conform to the duties and obligations of the role-positions they expect to occupy.

The contrast with anthropology and biology could hardly be more complete. Unlike anthropology, which celebrates the irreducible heterogeneity of cultures, sociology sees cultures as sharing much in common throughout the world (Brown 1991). In virtually every society, says sociology, youth are pressed to internalize the values of being trustworthy, loyal, helpful, friendly, courteous, kind, obedient, cheerful, thrifty, brave, clean, and reverent (famously captured by the Boy Scouts of America). In biology, values are collapsed into techniques, and the machinery of internalization is unrepresented.

Internalized norms are followed not because of their epistemic truth value, but because of their moral value. In the language of the BPC model, internalized norms are accepted not as instruments towards achieving other ends but rather as *arguments in the preference function that the individual maximizes*, or are *self-imposed constraints*. For example, individuals who have internalized the value of "speaking truthfully" will constrain themselves to do so even in some cases where the net payoff to speaking truthfully would otherwise be negative. Internalized norms are therefore *constitutive*, in the sense that an individual strives to live up to them *for their own sake*. Fairness, honesty, trustworthiness, and loyalty are ends, not means; and such fundamental human emotions as shame, guilt, pride, and empathy are deployed by the well-socialized individual to reinforce these prosocial values when tempted by the immediate pleasures of such "deadly sins" as anger, avarice, gluttony, and lust.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

The human responsiveness to socialization pressures represents perhaps the most powerful form of epigenetic transmission found in nature. In effect, *human preferences are programmable*, in the same sense that a computer can be programmed to perform a wide variety of tasks. This epigenetic flexibility, which is an emergent property of the complex human brain, in considerable part accounts for the stunning success of the species *Homo sapiens*. When people internalize a norm, the frequency of its occurrence in the population will be higher than if people follow the norm only instrumentally – that is, only when they perceive it to be in their material self-interest to do so. The increased incidence of altruistic pro-social behaviors permits humans to cooperate effectively in groups (Gintis et al. 2005a).

Given the abiding disarray in the behavioral sciences, it should not be surprising to find that socialization has no conceptual standing outside of sociology, anthropology, and social psychology, and that most behavioral scientists subsume it under the general category of "information transmission," which would make sense only if moral values expressed matters of fact, which they do not. Moreover, the socialization concept is incompatible with the assumption in economic theory that preferences are mostly, if not exclusively, self-regarding, given that social values commonly involve caring about fairness and the well-being of others Sociology, in turn, systematically ignores the limits to socialization (Pinker 2002; Tooby & Cosmides 1992) and supplies no theory of the emergence and abandonment of particular values, both of which depend in part on the contribution of values to fitness and well-being, as economic and biological theory would suggest (Gintis 2003a; 2003b). Moreover, there are often swift society-wide value changes that cannot be accounted for by socialization theory (Gintis 1975; Wrong 1961). When properly qualified, however, and appropriately related to the general theory of cultural evolution and strategic learning, socialization theory is considerably strengthened.

8. Game theory: The universal lexicon of life

In the BPC model, choices give rise to probability distributions over outcomes, the expected values of which are the payoffs to the choice from which they arose. Game theory extends this analysis to cases where there are multiple decision makers. In the language of game theory, *players* are endowed with a set of *strategies*, and they have certain *information* concerning the rules of the game, the nature of the other players, and their available strategies. Finally, for each combination of strategy choices by the players, the game specifies a distribution of *individual payoffs* to the players. Game theory predicts the behavior of the players by assuming each maximizes its preference function subject to its information, beliefs, and constraints (Kreps 1990).

Game theory is a logical extension of evolutionary theory. To see this, suppose there is only one replicator, deriving its nutrients and energy from nonliving sources (the sun, the Earth's core, amino acids produced by electrical discharge, and the like). The replicator population will then grow at a geometric rate, until it presses upon its environmental inputs. At that point, mutants that exploit the environment more efficiently will out-compete their less efficient conspecifics; and with input scarcity, mutants will emerge that "steal" from conspecifics who have amassed valuable resources. With the rapid growth of such mutant predators, their prey will mutate, thereby devising means of avoiding predation, and the predators will counter with their own novel predatory capacities. In this manner, strategic interaction is born from elemental evolutionary forces. It is only a conceptually short step from this point to cooperation and competition among cells in a multi-cellular body, among conspecifics who cooperate in social production, between males and females in a sexual species, between parents and offspring, and among groups competing for territorial control (Maynard Smith & Szathmary 1995/1997).

Historically, game theory emerged not from biological considerations but, rather, from the strategic concerns of combatants in World War II (Poundstone 1992; Von Neumann & Morgenstern 1944). This led to the widespread caricature of game theory as applicable only to static confrontations of rational self-regarding agents possessed of formidable reasoning and information-processing capacity. Developments within game theory in recent years, however, render this caricature inaccurate.

First, game theory has become the basic framework for modeling animal behavior (Alcock 1993; Krebs & Davies 1997a; Maynard Smith 1982), and thus has shed its static and hyperrationalistic character, in the form of *evolutionary game theory* (Gintis 2000c). The players in evolutionary game theory do not require the formidable information-processing capacities of the players in classical game theory, so disciplines that recognize that cognition is scarce and costly can make use of evolutionary game-theoretic models (Gigerenzer & Selten 2001; Gintis 2000c; Young 1998). Therefore, we may model individuals as considering only a restricted subset of strategies (Simon 1972; Winter 1971), and as using rule-of-thumb heuristics rather than maximization techniques (Gigerenzer & Selten 2001). Game theory is therefore a generalized schema that permits the precise framing of meaningful empirical assertions, but imposes no particular structure on the predicted behavior.

Second, evolutionary game theory has become a key to understanding the fundamental principles of evolutionary biology. Throughout much of the twentieth century, classical population biology did not employ a game-theoretic framework (Fisher 1930; Haldane 1932; Wright 1931). However, Moran (1964) showed that Fisher's Fundamental Theorem – which states that as long as there is positive genetic variance in a population, fitness increases over time – is false when more than one genetic locus is involved. Eshel and Feldman (1984) identified the problem with the population genetic model in its abstraction from mutation. But how do we attach a fitness value to a mutant? Eshel and Feldman (1984) suggested that payoffs be modeled game-theoretically on the phenotypic level and that a mutant gene be associated with a strategy in the resulting game. With this assumption, they showed that under some restrictive conditions, Fisher's Fundamental Theorem (Fisher 1930) could be restored. Their results have been generalized by Liberman (1988), Hammerstein and Selten (1994), Hammerstein (1996), Eshel et al. (1998), and others.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Third, the most natural setting for biological and social dynamics is game theoretic. Replicators (genetic and/or cultural) endow copies of themselves with a repertoire of strategic responses to environmental conditions, including information concerning the conditions under which each strategy is to be deployed in response to the character and density of competing replicators. Genetic replicators have been understood since the rediscovery of Mendel's laws in the early twentieth century. Cultural transmission also apparently occurs at the neuronal level in the brain, perhaps in part through the action of *mirror neurons*, which fire when either the individual performs a task or undergoes an experience, or when the individual observes another individual performing the same task or undergoing the same experience (Meltzhoff & Decety 2003; Rizzolatti et al. 2002; Williams et al. 2001). Mutations include replacement of strategies by modified strategies; and the "survival of the fittest" dynamic (formally called a *replicator dynamic*) ensures that replicators with more successful strategies replace those with less successful (Taylor & Jonker 1978).

Fourth, behavioral game theorists, who use game theory to collect experimental data concerning strategic interaction, now widely recognize that in many social interactions, individuals are not self-regarding. Rather, they often care about the payoffs to and intentions of other players, and they will sacrifice to uphold personal standards of honesty and decency (Fehr & Gächter 2002; Gintis et al. 2005a; Gneezy 2005; Wood 2003). Moreover, humans care about power, self-esteem, and behaving morally (Bowles & Gintis 2005a; Gintis 2003a; Wood 2003). Because the rational actor model treats action as instrumental towards achieving rewards, it is often inferred that action itself cannot have reward value. This is an unwarranted inference. For instance, the rational actor model can be used to explain collective action (Olson 1965), since individuals may place positive value on the process of acquisition (e.g., "fighting for one's rights"), and they can value punishing those who refuse to join in the collective action (Moore 1978; Wood 2003). Indeed, contemporary experimental work indicates that one can apply standard choice theory, including the derivation of demand curves, plotting concave indifference curves, and finding price elasticities, for such preferences as charitable giving and punitive retribution (Andreoni & Miller 2002).

As a result of its maturation over the past quarter century, game theory is well positioned to serve as a bridge across the behavioral sciences, providing both a lexicon for communicating across fields with distinct and incompatible conceptual systems and a theoretical tool for formulating a model of human choice that can serve all the behavioral disciplines.

9. Some misconceptions concerning the BPC model and game theory

Many behavioral scientists reject the BPC model and game theory on the basis of one or more of the following arguments. In each case, I shall indicate why the objection is not compelling.

9.1. Individuals are only boundedly rational

Perhaps the most pervasive critique of the BPC model is that put forward by Herbert Simon (1982), holding that because information processing is costly and humans have finite information-processing capacity, individuals *satisfice* rather than *maximize*, and hence are only *boundedly rational*. There is much substance to this view, including the importance of taking into account information-processing costs and limited information in modeling choice behavior and recognizing that the decision on how much information to collect depends on unanalyzed subjective priors at some level (Heiner 1983; Winter 1971). Indeed, from basic information theory and the Second Law of Thermodynamics it follows that *all rationality is bounded*. However, the popular message taken from Simon's work is that we should reject the BPC model. For example, the mathematical psychologist D. H. Krantz (1991) asserts, "The normative assumption that individuals *should* maximize *some* quantity may be wrong. . . . People do and should act as *problem solvers*, not *maximizers*." This is incorrect. As we have seen, as long as individuals have consistent preferences, they can be modeled as maximizing an objective function.

Of course, if there is a precise objective (e.g., solve the problem with an exogenously given degree of accuracy), then the information contained in knowledge of preference consistency may be ignored. But, once the degree of acceptability is treated as endogenous, multiple objectives compete (e.g., cost and accuracy), and the BPC model cannot be ignored. This point is lost on even such capable researchers as Gigerenzer and Selten (2001), who reject the "optimization subject to constraints" method on the grounds that individuals do not in fact solve optimization problems. However, just as billiards players do not solve differential equations in choosing their shots, so decision makers do not solve Lagrangian equations, even though in both cases we may use optimization models to describe their behavior.

9.2. Decision makers are not consistent

It is widely argued that in many situations of extreme importance, choice consistency fails, so preferences are not maximized. These cases include time inconsistency, in which individuals have very high short-term discount rates and much lower long-term discount rates (Ainslie 1975; Herrnstein 1961; Laibson 1997). As a result, people lack the will-power to sacrifice present pleasures for future well-being. This leads to such well-known behavioral problems as unsafe sex, crime, substance abuse, procrastination, under-saving, and obesity. It is therefore held that these phenomena of great public policy importance are irrational and cannot be treated with the BPC model.

When the choice space for time preference consists of pairs of the form *(reward, delay until reward materializes)*, then preferences are indeed time inconsistent. The long-term discount rate can be estimated empirically at about 3% per year (Huang & Litzenberger 1988; Rogers 1994), but short-term discount rates are often of an order of magnitude greater than this (Laibson 1997). Animal studies find rates that are even several orders of magnitude higher (Stephens et al. 2002). Consonant with these findings, sociological theory stresses that *impulse control* – learning to favor long-term over short-term gains – is a major component in the socialization of youth (Grusec & Kuczynski 1997; Power & Chapieski 1986).

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

However, suppose we expand the choice space to consist of triples of the form *(reward, current time, time when reward accrues)*, so that, for instance, $(\pi_1, t_1, s_1) > (\pi_2, t_2, s_2)$ means that the individual prefers to be at time t_1 facing a reward π_1 delivered at time s_1 to being at time t_2 facing a reward π_2 delivered at time s_2. Then the observed behavior of individuals with discount rates that decline with the delay becomes choice consistent, and there are two simple models that are roughly consistent with the available evidence (and differ only marginally from each other): hyperbolic and quasi-hyperbolic discounting (Ahlbrecht & Weber 1995; Ainslie & Haslam 1992; Fishburn & Rubinstein 1982; Laibson 1997). The resulting BPC models allow for sophisticated and compelling economic analyses of policy alternatives (Laibson et al. 2004).

Other observed instances of prima facie choice inconsistency can be handled in a similar fashion. For example, in experimental settings, individuals exhibit status quo bias, loss aversion, and regret – all of which imply inconsistent choices (Kahneman & Tversky 1979; Sugden 1993a). In each case, however, choices become consistent by a simple redefinition of the appropriate choice space. Kahneman and Tversky's "prospect theory," which models status quo bias and loss aversion, is precisely of this form. Gintis (in press b) has shown that this phenomenon has an evolutionary basis in territoriality in animals and pre-institutional property rights in humans.

There remains perhaps the most widely recognized example of inconsistency, that of preference reversal in the choice of lotteries. Lichtenstein and Slovic (1971) were the first to find that in many cases, individuals who prefer lottery A to lottery B are nevertheless willing to take less money for A than for B. Reporting this to economists several years later, Grether and Plott (1979) asserted, "A body of data and theory has been developed . . . [that] are simply inconsistent with preference theory" (p. 623). These preference reversals were explained several years later by Tversky et al. (1990) as a bias towards the higher probability of winning in lottery choice and towards the higher the maximum amount of winnings in monetary valuation. If this were true for lotteries in general, it might compromise the BPC model.[11] However, the phenomenon has been documented only when the lottery pairs A and B are so close in expected value that one needs a calculator (or a quick mind) to determine which would be preferred by an expected value maximizer. For example, in Grether and Plott (1979) the average difference between expected values of comparison pairs was 2.51% (calculated from their Table 2, p. 629). The corresponding figure for Tversky et al. (1990) was 13.01%. When the choices are so close to indifference, it is not surprising that inappropriate cues are relied upon to determine choice. Moreover, Berg et al. (2005) have shown that when analysis is limited to studies that have truth-revealing incentives, preference reversals are well described by a model of maximization with error.

Another source of inconsistency is that observed preferences may not lead to the well-being, or even the immediate pleasure, of the decision maker. For example, fatty foods and tobacco injure health yet are highly prized; addicts often say they get no pleasure from consuming their drug of choice but are driven by an inner compulsion to consume; and individuals with obsessive-compulsive disorders repeatedly perform actions that they know are irrational and harmful. More generally, behaviors resulting from excessively high short-term discount rates, discussed above, are likely to lead to a divergence of choice and welfare.

However, the BPC model is based on the premise that choices are consistent, not that choices are highly correlated with welfare. Drug addiction, unsafe sex, unhealthy diet, and other individually welfare-reducing behaviors can be analyzed with the BPC model, although in such cases preferences and welfare may diverge. I have argued that we can expect the BPC to hold because, on an evolutionary time scale, brain characteristics will be selected according to their capacity to contribute to the fitness of their bearers. But, fitness cannot be equated with well-being in any creature. Humans, in particular, live in an environment so dramatically different from that in which our preferences evolved that it seems to be miraculous that we are as capable as we are of achieving high levels of individual well-being. For instance, in virtually all known cases, fertility increases with per capita material wealth in a society up to a certain point and then decreases. This is known as the *demographic transition*, and it accounts for our capacity to take out increased technological power in the form of consumption and leisure rather than increased numbers of offspring (Borgerhoff Mulder 1998). No other known creature behaves in this fashion. Thus, our preference predispositions have not "caught up" with our current environment and, given especially the demographic transition and our excessive present-orientation, they may never catch up (Akerlof 1991; Elster 1979; O'Donoghue & Rabin 2001).

9.3. Addiction contradicts the BPC model

Substance abuse is of great contemporary social importance and is held clearly to violate the notion of rational behavior. Substance abusers are often exhibited as prime examples of time inconsistency and the discrepancy between choice and well-being. But, as discussed above, these characteristics do not invalidate the use of the BPC model. More telling, perhaps, is the fact that even draconian increases in the penalties for illicit substance use do not lead to the abandonment of illegal substances. In the United States, for example, the "war on drugs" has continued for several decades; yet, despite the dramatic increase in the prison population, it has not effectively curbed the illicit behavior. Since the hallmark of the rational actor model is that individuals trade off among desired goals, the lack of responsiveness of substance abuse to dramatically increased penalties has led many researchers to reject the BPC.

The target of much of the criticism of the BPC approach to substance abuse is the work of economist Gary Becker and his associates; in particular, the seminal paper on addiction by Becker and Murphy (1988). Many aspects of the Becker-Murphy "rational addiction" model are accurate, however; and subsequent empirical research has validated the notion that illicit drugs respond to market forces much as any marketed good or service. For instance, Saffer and Chaloupka (1999) estimated the price elasticities of heroin and cocaine using a sample of 49,802 individuals from the National Household Survey

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

of Drug Abuse. The price elasticities for heroin and cocaine were about 1.70 and 0.96, respectively, which are quite high. Using these figures, the authors estimate that the lower prices flowing from the legalization of these drugs would lead to an increase of about 100% and 50% in the quantities of heroin and cocaine consumed, respectively.

How does this square with the observation that draconian punishments do not squelch the demand altogether? Gruber and Koszegi (2001) explain this by presenting evidence that drug users exhibit the commitment and self-control problems that are typical of time-inconsistent individuals, for whom the possible future penalties have highly attenuated deterrent value in the present. Nevertheless, allowing for this attenuated value, sophisticated economic analysis of the sort developed by Becker et al. (1994) can be deployed for policy purposes. Moreover, this analytical and quantitative analysis harmonizes with the finding that, along with raising the price of cigarettes, the most effective way to reduce the incidence of smoking is to raise its immediate personal costs, such as being socially stigmatized, being banned from smoking in public buildings, and being considered impolite, given the well-known externalities associated with second-hand smoke (Brigden & De Beyer 2003).

9.4. Positing exotic tastes explains nothing

Some have argued that broadening the rational actor model beyond its traditional form in neoclassical economics runs the risk of developing unverifiable and post hoc theories, as our ability to theorize outpaces our ability to test theories. Indeed, the folklore among economists dating back at least to Becker and Stigler (1977) is that "you can always explain any bizarre behavior by assuming sufficiently exotic preferences."

This critique was telling before researchers had the capability of actually measuring preferences and testing the cogency of models with nonstandard preferences (i.e., preferences concerning things other than marketable commodities, forms of labor, and leisure). However, behavioral game theory now provides the methodological instruments for devising experimental techniques that allow us to estimate preferences with some degree of accuracy (Camerer 2003; Gintis 2000c). Moreover, we often find that the appropriate experimental design variations can generate novel data allowing us to distinguish among models that are equally powerful in explaining the existing data (Kiyonari et al. 2000; Tversky & Kahneman 1981). Finally, because behavioral game-theoretic predictions can be systematically tested, the results can be replicated by different laboratories (Plott 1979; Sally 1995; V. Smith 1982), and models with very few nonstandard preference parameters (examples of which are provided in Sect. 10 below) can be used to explain a variety of observed choice behavior.

9.5. Decisions are sensitive to framing bias

The BPC model assumes that individuals have stable preferences and beliefs that are functions of the individual's personality and current needs. Yet, in many cases laboratory experiments show that individuals can be induced to make choices over payoffs based on subtle or obvious cues that ostensibly do not affect the value of the payoffs to the decision maker. For example, if a subject's partner in an experimental game is described an "opponent," or if the game itself is described as a "bargaining game," then the subject may make very different choices than when the partner is described as a "teammate," or if the game is described as a "community participation game." Similarly, a subject in an experimental game may reject an offer if made by his bargaining partner, but accept the same offer if made by the random draw of a computer on behalf of the proposer (Blount 1995).

Sensitive to this critique, experimenters in the early years of behavioral game theory attempted to minimize the possibility of framing effects by rendering the language in which a decision problem or strategic interaction was described as abstract and unemotive as possible. It is now widely recognized that framing effects cannot be avoided, because abstraction and lack of real-world reference are themselves a frame rather than an absence thereof. A more productive way to deal with framing is to make the frame a part of the specification of the experiment itself. Varying the frame systematically will uncover the effect of the frame on the choices of the subjects, and by inference, on their beliefs and preferences.

We do not have a complete understanding of framing, but we do know enough to assert that its existence does not undermine the BPC model. If subjects care only about the "official" payoffs in a game, and if framing does not affect the beliefs of the subjects as to what other subjects will do, then framing could not affect behavior in the BPC model. But, subjects generally do care about fairness, reciprocity, and justice, as well as about the game's official payoffs; when confronted with a novel social setting in the laboratory, subjects must first decide what moral values to apply to the situation by *mapping the game onto some sphere of everyday life* to which they are accustomed. The verbal and other cues provided by experimenters are the clues that subjects use to "locate" the interaction in their social space, so that moral principles can be properly applied to the novel situation. Moreover, framing instruments such as calling subjects "partners" rather than "opponents" in describing the game can increase cooperation, because *strong reciprocators* (Gintis 2000d), who prefer to cooperate if others do the same, may increase their assessment of the probability that others will cooperate (see sect. 10), if given the "partner" as opposed to the "opponent" cue. In sum, framing is in fact an ineluctable part of the BPC model, properly construed.

9.6. People are faulty logicians

The BPC model permits us to infer the beliefs and preferences of individuals from their choices under varying constraints. Such inferences are valid, however, only if individuals can intelligently vary their behavior in response to novel conditions. It is common for behavioral scientists who reject the BPC model to explain an observed behavior as the result of an error or confusion on the part of the individual. But the BPC model is less tolerant of such explanations, if individuals are reasonably well informed and the choice setting is reasonably transparent and easily analyzable.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Evidence from experimental psychology over the past 40 years has led some psychologists to doubt the capacity of individuals to reason sufficiently accurately to warrant the BPC presumption of subject intelligence. For example, in one well-known experiment performed by Tversky and Kahneman (1983), a young woman, Linda, is described as politically active in college and highly intelligent; then the subject is asked which of the following two statements is more likely: "Linda is a bank teller" or "Linda is a bank teller and is active in the feminist movement." Many subjects rate the second statement more likely, despite the fact that elementary probability theory asserts that if p implies q, then p cannot be more likely than q. Because the second statement implies the first, it cannot be more likely than the first.

I personally know many people (though not scientists) who give this "incorrect" answer, and I never have observed these individuals making simple logical errors in daily life. Indeed, in the literature on the "Linda problem," several alternatives to faulty reasoning have been offered. One highly compelling alternative is based on the notion that in normal conversation, a listener assumes that any information provided by the speaker is relevant to the speaker's message (Grice 1975). Applied to this case, the norms of conversation lead the subject to believe that the experimenter wants Linda's politically active past to be taken adequately into account (Hilton 1995; Wetherick 1995). Moreover, the meaning of such terms as "more likely" or "higher probability" are vigorously disputed even in the theoretical literature, and hence are likely to have a different meaning for the average subject versus for the expert. For example, if I were given two piles of identity folders and asked to search through them to find the one belonging to Linda, and one of the piles was "all bank tellers" while the other was "all bank tellers who are active in the feminist movement," I would surely look through the latter (doubtless much smaller) pile first, even though I am well aware that there is a "higher probability" that Linda's folder is in the former pile rather than the latter one.

More generally, subjects may appear irrational because basic terms have different meanings in propositional logic versus in everyday logical inference. For example, "if p then q" is true in formal logic except when p is true and q is false. In everyday usage "if p then q" may be interpreted as a material implication, in which there is something about p that causes q to be the case. In particular, in material logic "p implies q" means "p is true and this situation causes q to be true." Similarly, "if France is in Africa, then Paris is in Europe" is true in propositional logic, but false as a material implication. Part of the problem is also that individuals without extensive academic training simply lack the expertise to follow complex chains of logic, so psychology experiments often exhibit a high level of *performance error* (Cohen 1981; see sect. 11). For instance, suppose Pat and Kim live in a certain town where all men have beards and all women wear dresses. Then the following can be shown to be true in propositional logic: "Either if Pat is a man, then Kim wears a dress or if Kim is a woman, then Pat has a beard." It is quite hard to see why this is formally true, and it is not true if the implications are material. Finally, the logical meaning of "if p then q" can be context dependent. For example, "if you eat dinner (p), you may go out to play (q)" formally means "you may go out to play (q) only if you eat dinner (p)."

We may apply this insight to an important strand of experimental psychology that purports to have shown that subjects systematically deviate from simple principles of logical reasoning. In a widely replicated study, Wason (1966) showed subjects cards each of which had a numeral 1 or 2 on one side and a letter *A* or *B* on the other, and stated the following rule: "A card with a vowel on one side must have an odd number on the other." The experimenter then showed each subject four cards – one showing 1, one showing 2, one showing *A*, and one showing *B* – and asked the subject which cards must be turned over to check whether the rule was followed. Typically, only about 15% of college students pointed out the correct cards (*A* and 2). Subsequent research showed that when the problem is posed in more concrete terms, such as "any person drinking beer must be over 18," the correct response rate increases considerably (Cheng & Holyoak 1985; Cosmides 1989; Shafir & LeBoeuf 2002; Stanovich 1999). This accords with the observation that most individuals do not appear to have difficulty making and understanding logical arguments in everyday life.

9.7. People are poor statistical decision makers

Just as the rational actor model began to take hold in the mid-twentieth century, vigorous empirical objections began to surface. The first was Allais (1953), who described cases in which subjects exhibited clear inconsistency in choosing among simple lotteries .It has been shown that Allais' examples can be explained by regret theory (Bell 1982; Loomes & Sugden 1982), which can be represented by consistent choices over pairs of lotteries (Sugden 1993a).

Close behind Allais came the famous Ellsberg Paradox (Ellsberg 1961), which can be shown to violate the most basic axioms of choice under uncertainty. Consider two urns. Urn A has 51 red balls and 49 white balls. Urn B also has 100 red and white balls, but the fraction of red balls is unknown. Subjects are asked to choose in two situations. In each, the experimenter draws one ball from each urn but the two balls remain hidden from the subject's sight. In the first situation, the subject can choose the ball that was drawn from urn A or urn B, and if the ball is red, the subject wins $10. In the second situation, the subject again can choose the ball drawn from urn A or urn B, and if the ball is white, the subject wins $10. Many subjects choose the ball drawn from urn A in both situations. This obviously violates the expected utility principle, no matter what probability the subject places on the likelihood that the ball from urn B is white.

It is easy to see why unsophisticated subjects make this error: Urn B seems to be *riskier* than urn A, because we know the probabilities in A but not in B. It takes a relatively sophisticated probabilistic argument – one that no human being ever made or could have made (to our knowledge) prior to the modern era – to see that in fact, in this case, uncertainty does not lead to increased risk. Indeed, most intelligent subjects who make the Ellsberg error will be convinced, when presented with the logical analysis, to modify their choices without modifying their preferences. In cases like this, we speak of performance

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

error, whereas in cases such as the Allais Paradox, even the most highly sophisticated subject will need to change his choice unless convinced to change his preference ordering.

Numerous experiments document that many people have beliefs concerning probabilistic events that are without scientific foundation, and which will likely lead them to sustain losses if acted upon. For example, virtually every enthusiast believes that athletes in competitive sports run "hot and cold," although this has never been substantiated empirically. In basketball, when a player has a "hot hand," he is preferentially allowed to shoot again, and when he has a "cold hand," he is often taken out of the game. I have yet to meet a basketball fan who does not believe in this phenomenon. Yet Gilovich et al. (1985) have shown, on the basis of a statistical analysis using professional basketball data, that the "hot/cold hand" does not exist.[12] This is but one instance of the general rule that our brains often lead us to perceive a pattern when faced with purely random data. In the same vein, I have talked to professional stock traders who believe, on the basis of direct observation of stock volatility, that stocks follow certain laws of inertia and elasticity that cannot be found through a statistical analysis of the data. Another example of this type is the "gambler's fallacy," which is that in a fair game, the appearance of one outcome several times in a row renders that outcome less likely in the next several plays of the game. Those who believe this cannot be dissuaded by scientific evidence. Many who believe in the "Law of Small Numbers," which says that a small sample from a large population will have the same distribution of characteristics as is in the population (Tversky & Kahneman 1971), simply cannot be dissuaded either by logical reasoning or by presentation of empirical evidence.

We are indebted to Daniel Kahneman, Amos Tversky, and their colleagues for a series of brilliant papers, beginning in the early 1970s, documenting the various errors that intelligent subjects commit in dealing with probabilistic decision making. Subjects systematically underweight base rate information in favor of salient and personal examples; they reverse lottery choices when the same lottery is described by emphasizing probabilities rather than monetary payoffs, or when described in terms of losses from a high baseline as opposed to gains from a low baseline; and they treat proactive decisions differently from passive decisions even when the outcomes are exactly the same and when outcomes are described in terms of probabilities as opposed to frequencies (Kahneman et al. 1982; Kahneman & Tversky 2000).

These findings are important for understanding human decision making and for formulating effective social policy mechanisms where complex statistical decisions must be made. However, these findings are not a threat to the BPC model (Gigerenzer & Selten 2001). They are simply performance errors in the form of incorrect beliefs as to how payoffs can be maximized.[13]

Statistical decision theory did not exist until recently. Before the contributions of Bernoulli, Savage, von Neumann, and other experts, no creature on Earth knew how to value a lottery. It takes years of study to feel at home with the laws of probability. Moreover, it is costly, in terms of time and effort, to apply these laws even if we know them. Of course, if the stakes are high enough, it is worthwhile to go to the effort, or engage an expert who will do it for you. But generally, we apply a set of heuristics that more or less get the job done (Gigerenzer & Selten 2001). Among the most prominent heuristics is simply *imitation*: decide what class of phenomenon is involved, find out what people "normally do" in that situation, and do it. If there is some mechanism leading to the survival and growth of relatively successful behaviors and if the problem in question recurs with sufficient regularity, the choice-theoretic solution will describe the winner of a dynamic social process of trial, error, and replication through imitation.

9.8. Classical game theory misunderstands rationality

Game theory predicts that rational agents will play Nash equilibria. Since my proposed framework includes both game theory and rational agents, I must address the fact that in important cases, the game-theoretic prediction is ostensibly falsified by the empirical evidence. The majority of examples of this kind arise from the assumption that individuals are self-regarding, which can be dropped without violating the principles of game theory. Game theory also offers solutions to problems of cooperation and coordination which are never found in real life; but in this case, the reason is that the game theorists assume perfect information, the absence of errors, the use of solution concepts that lack plausible dynamical stability properties, or other artifices without which the proposed solution would not work (Gintis 2005b). However, in many cases, rational agents simply do not play Nash equilibria at all under plausible conditions.

Consider, for example, the centipede game, depicted in Figure 1 (Binmore 1987; Rosenthal 1981). It is easy to show that this game has only one Nash payoff structure, in which player one defects on round one. However, when people actually play this game, they generally cooperate until the last few rounds (McKelvey & Palfrey 1992). Game theorists are quick to call such cooperation "irrational." For instance, Reinhard Selten (himself a strong supporter of "bounded rationality") considers any move other than immediate defection a "failure to behave according to one's rational insights" (Selten 1993, p. 133). This opinion is due to the fact that this is the unique Nash equilibrium to the game, it does not involve the use of mixed strategies, and it can be derived from backward induction. However, as the professional literature of the past two decades makes abundantly clear, it is simply not true that rational agents must use backward induction. Rather, the most that rationality can ensure is *rationalizability* (Bernheim 1984; Pearce 1984), which in the case of the centipede game includes any pair of actions, except for cooperation on a player's final move.

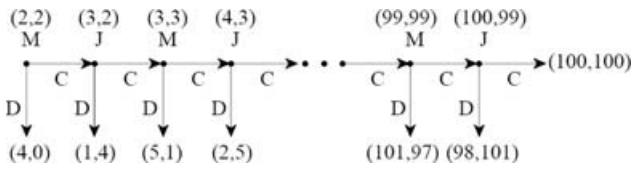


Figure 1. The Hundred Round Centipede Game illustrates the fallacy of holding that "rational" agents must use backward induction in their strategic interactions.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Indeed, the epistemic conditions under which it is reasonable to assert that rational agents will play a Nash equilibrium are plausible in only the simplest cases (Aumann & Brandenburger 1995).

Another way to approach this issue is to begin by simply endowing each player with a BPC structure and defining each player's *type* to be the round on which the player would first defect, assuming this round is reached. The belief system of each player is then a subjective probability distribution over the type of his or her opponent. It is clear that if players attempt to maximize their payoffs subject to this probability distribution, many different actions can result. Indeed, when people play this game, they generally cooperate at least until the final few rounds. This, moreover, is an eminently correct solution to the problem, and much more lucrative that the Nash equilibrium. Of course, one could argue that both players must have the *same* subjective probability distribution (this is called the *common priors* assumption) – in which case (assuming common priors are common knowledge) there is only one equilibrium, the Nash equilibrium. But, it is hardly plausible to assume two players have the same subjective probability distribution over the types of their opponents without giving a mechanism that would produce this result.[14] In a famous paper, Nobel prize winning economist John Harsanyi (1967) argued that common priors follow from the assumption that individuals are rational. But, this argument depends on a notion of rationality that goes far beyond choice consistency, and it has not received empirical support (Kurz 1997).

In real-world applications of game theory, I conclude, we must have plausible grounds for believing that the equilibrium concept used is appropriate. Simply assuming that rationality implies Nash equilibrium, as is the case in classical game theory, is generally inappropriate. Evolutionary game theory restores the centrality of the Nash equilibrium concept, because stable equilibria of the replicator dynamic (and related "monotone" dynamics) are necessarily Nash equilibria. Moreover, the examples given in the next section are restricted to games that are sufficiently simple that the sorts of anomalies discussed above are not present, and the Nash equilibrium criterion is appropriate.

10. Behavioral game theory and other-regarding preferences

Contemporary biological theory maintains that cooperation can be sustained by means of *inclusive fitness*, or cooperation among kin (Hamilton 1963) and by individual self-interest in the form of *reciprocal altruism* (Trivers 1971). Reciprocal altruism occurs when an individual helps another individual, at a fitness cost to itself, contingent on the beneficiary returning the favor in a future period. The explanatory power of inclusive fitness theory and reciprocal altruism convinced a generation of biologists that what appears to be altruism – personal sacrifice on behalf of others – is really just long-run genetic self-interest.[15] Combined with a vigorous critique of group selection (Dawkins 1976; Maynard Smith 1976; Williams 1966), a generation of biologists became convinced that true altruism – one organism sacrificing fitness on behalf of the fitness of an unrelated other – was virtually unknown, even in the case of *Homo sapiens*.

That human nature is selfish was touted as a central implication of rigorous biological modeling. In *The Selfish Gene*, for example, Richard Dawkins asserts that "We are survival machines – robot vehicles blindly programmed to preserve the selfish molecules known as genes.... Let us try to teach generosity and altruism, because we are born selfish" (Dawkins 1976, p. 7). Similarly, in *The Biology of Moral Systems*, R. D. Alexander asserts that "ethics, morality, human conduct, and the human psyche are to be understood only if societies are seen as collections of individuals seeking their own self-interest" (Alexander 1987, p. 3). More poetically, Michael Ghiselin writes, "No hint of genuine charity ameliorates our vision of society, once sentimentalism has been laid aside. What passes for cooperation turns out to be a mixture of opportunism and exploitation.... Scratch an altruist, and watch a hypocrite bleed" (Ghiselin 1974, p. 3).

In economics, the notion that enlightened self-interest allows individuals to cooperate in large groups goes back to Bernard Mandeville's "private vices, public virtues" (Mandeville 1705/1924) and Adam Smith's "invisible hand" (Smith 1759/2000). Full analytical development of this idea awaited the twentieth-century development of general equilibrium theory (Arrow & Debreu 1954; Arrow & Hahn 1971) and the theory of repeated games (Axelrod & Hamilton 1981; Fudenberg & Maskin 1986).

By contrast, sociological, anthropological, and social psychological theory generally explain that human cooperation is predicated upon affiliative behaviors among group members, each of whom is prepared to sacrifice a modicum of personal well-being to advance the group's collective goals. The vicious attack on "sociobiology" (Segerstrale 2001) and the widespread rejection of *Homo economicus* in the "soft" social sciences (DiMaggio 1994; Etzioni 1985; Hirsch et al. 1990) is due, in part, to this clash of basic explanatory principles.

Behavioral game theory assumes the BPC model, and it subjects individuals to strategic settings, such that their behavior reveals their underlying preferences. This controlled setting allows us to adjudicate between these contrasting models. One behavioral regularity that has been found thereby is *strong reciprocity*, which is a predisposition to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when it is implausible to expect that these costs will be repaid. Strong reciprocity is other-regarding, as a strong reciprocator's behavior reflects a preference to cooperate with other cooperators and to punish non-cooperators, even when these actions are personally costly.

The result of the laboratory and field research on strong reciprocity is that humans indeed often behave in ways that have traditionally been affirmed in sociological theory and denied in biology and economics (Andreoni 1995; Fehr & Gächter 2000; 2002; Fehr et al. 1997; 1998; Gächter & Fehr 1999; Henrich et al. 2005; Ostrom et al. 1992). Moreover, it is probable that this other-regarding behavior is a prerequisite for cooperation in large groups of non-kin, since the theoretical models of cooperation in large groups of self-regarding nonkin in biology and economics do not apply to some important and frequently observed forms of human cooperation (Boyd & Richerson 1992; Gintis 2005b).

Another form of prosocial behavior conflicting with the maximization of personal material gain is that of

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

maintaining such *character virtues* as honesty and promise-keeping, even when there is no chance of being penalized for unvirtuous behavior. An example of such behavior is reported by Gneezy (2005), who studied 450 undergraduate participants that were paired off to play three games of the following form: Player One would be shown two pairs of payoffs, A:(x, y) and B:(z, w) where x, y, z, and w are amounts of money with $x < z$ and $y > w$. Player One could then say to Player Two, who could not see the amounts of money, either "Option A will earn you more money than option B," or "Option B will earn you more money than option A," The first game was A:(5, 6) versus B:(6, 5) so Player One could gain 1 by lying and being believed, while imposing a cost of 1 on Player Two. The second game was A:(5, 15) versus B:(6, 5) so Player One could gain 10 by lying and being believed, while still imposing a cost of 1 on Player Two. The third game was A:(5, 15) versus B:(15, 5) so Player One could gain 10 by lying and being believed, while imposing a cost of 10 on Player Two. Before starting play, Gneezy asked all the Player One's whether they expected their advice to be followed, inducing honest responses by promising to reward subjects whose guesses were correct. He found that 82% of the Player Ones expected their advice to be followed (the actual number was 78%).

It follows from the Player One expectations that if they were self-regarding, they would always lie and recommend B to Player Two. In fact, in Game 2, where being deceived was very costly to Player Two and the gain to deceiving was small for Player One, only 17% of subjects lied. In Game 1, where Player Two's cost of being deceived to was only 1 but the gain to Player One of deceiving Player 2 was the same as in Game 2, 36% lied. In other words, subjects were loathe to lie, but considerably more so when it was costly to their partner. In Game 3, where the gain to deceiving was large for Player One, and equal to the loss from being deceived to Player Two, fully 52% lied.

This shows that many subjects are willing to sacrifice material gain to avoid lying in a one-shot, anonymous interaction, their willingness to lie increasing with an increased cost of truth-telling to themselves, and decreasing with an increase in their partner's cost of begin deceived. Similar results were found by Boles et al. (2000) and Charness and Dufwenberg (2004). In addition, Gunnthorsdottir et al. (2002) and Burks et al. (2003) have shown that a social-psychological measure of "Machiavellianism" predicts which subjects are likely to be trustworthy and trusting.

11. Beliefs: The weak link in the BPC model

In the simplest formulation of the rational actor model, beliefs do not explicitly appear (Savage 1954). In the real world, however, the probabilities of various outcomes in a lottery are rarely objectively known, and hence must generally be subjectively constructed as part of an individual's belief system. Anscombe and Aumann (1963) extended the Savage model to preferences over bundles consisting of "states of the world" and payoff bundles, and showed that if certain consistency axioms hold, the individual could be modeled as maximizing subject to a set of subjective probabilities (beliefs) over states. Were these axioms universally plausible, beliefs could be derived in the same way as preferences are derived. However, at least one of these axioms, the so-called *state-independence axiom*, which posits that preferences over payoffs are independent of the states in which they occur, is generally not plausible.

It follows that beliefs are the underdeveloped member of the BPC trilogy. Except for Bayes' rule (Gintis 2000c, Ch. 17), there is no compelling analytical theory of how a rational agent acquires and updates beliefs, although there are many partial theories (Boyer 2001; Jaynes 2003; Kuhn 1962; Polya 1990).

Beliefs enter the decision process in several potential ways. First, individuals may not have perfect knowledge concerning how their choices affect their welfare. This is most likely to be the case in an unfamiliar setting, of which the experimental laboratory is often a perfect example. In such cases, when forced to choose, individuals "construct" their preferences on the spot by forming beliefs based on whatever partial information is present at the time of choice (Slovic 1995). Understanding this process of belief formation is a demanding research task.

Second, often the actual actions $a \in A$ available to an individual will differ from the actual payoffs $\pi \in \Pi$ that appear in the individual's preference function. The mapping β: $A \rightarrow \Pi$ the individual deploys to maximize payoff is a belief system concerning objective reality, and it can differ from the correct mapping β^*: $A \rightarrow \Pi$. For example, a gambler may want to maximize expected winnings but may believe in the erroneous Law of Small Numbers (Rabin 2002). Errors of this type include the performance errors discussed in section 9.

Third, there is considerable evidence that beliefs directly affect well-being, so individuals may alter their beliefs as part of their optimization program. Self-serving beliefs, unrealistic expectations, and projection of one's own preferences on others are important examples. The trade-off here is that erroneous beliefs may add to well-being, but acting on these beliefs may lower other payoffs (Benabou & Tirole 2002; Bodner & Prelec 2002).

12. Conclusion

Each of the behavioral disciplines contributes strongly to understanding human behavior. Taken separately and at face value, however, they offer partial, conflicting, and incompatible models. From a scientific point of view, it is scandalous that this situation was tolerated throughout most of the twentieth century. Fortunately, there is currently a strong current of unification, based on both mathematical models and common methodological principles for gathering empirical data on human behavior and human nature.

The true power of each discipline's contribution to knowledge will only appear when suitably qualified and deepened by the contribution of the others. For example, the economist's model of rational choice behavior must be qualified by a biological appreciation that preference consistency is the result of strong evolutionary forces, and that where such forces are absent, consistency may be imperfect. Moreover, the notion that preferences are purely self-regarding must be abandoned. For a second example, the sociologist's notion of internalization of norms must be thoroughly integrated into behavioral theory, which must recognize that the ease with which

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

diverse values can be internalized depends on human nature (Pinker 2002; Tooby & Cosmides 1992). The rate at which values are acquired and abandoned depends on their contribution to fitness and well-being (Gintis 2003a; 2003b) – there are often rapid, society-wide value changes that cannot be accounted for by socialization theory (Gintis 1975; Wrong 1961).

Disciplinary boundaries in the behavioral sciences have been determined historically, rather than conforming to some consistent scientific logic. Perhaps for the first time, we are in a position to rectify this situation. We must recognize evolutionary theory (covering both genetic and cultural evolution) as the integrating principle of behavioral science. Moreover, if the BPC model is broadened to encompass other-regarding preferences and a cogent theory of belief formation and change is developed, game theory becomes capable of modeling all aspects of decision making, including those normally considered "sociological" or "anthropological," which in turn is most naturally the central organizing principle of psychology.

ACKNOWLEDGMENTS

I would like to thank George Ainslie, Rob Boyd, Dov Cohen, Ernst Fehr, Barbara Finlay, Ernst Fehr, Thomas Getty, Joe Henrich, Daniel Kahneman, Laurent Keller, Joachim Krueger, Larry Samuelson, Robert Trivers, and, especially, Marc Hauser and anonymous referees of this journal for helpful comments, and the John D. and Catherine T. MacArthur Foundation for financial support.

NOTES

1. Biology straddles the natural and behavioral sciences. We include biological models of animal (including human) behavior, as well as the physiological bases of behavior, in the behavioral sciences.

2. The last serious attempt at developing an analytical framework for the unification of the behavioral sciences was by Parsons and Shils (1951). A more recent call for unity is made by Wilson (1998), who does not supply the unifying principles.

3. A core contribution of political science, the concept of power, is absent from economic theory, yet interacts strongly with basic economic principles (Bowles & Gintis 2000). Lack of space prevents me from expanding on this important theme.

4. Throughout this target article, fitness refers to inclusive fitness (Hamilton 1963).

5. I use the term "self-regarding" to avoid the confusion that results from using the more common term "self-interested" when the individual prefers outcomes that benefit others. For example, if I prefer to give gifts to others, my behavior in doing so may be selfish (it may give me pleasure), but it is certainly not self-regarding (the pleasure comes from another individual's gains).

6. Throughout this paper, I generalize concerning the nature of disciplines (e.g., psychology, economics) and subdisciplines (e.g., cognitive psychology, neoclassical economics) that I believe accurately depict their broad nature, their common core, and the way they are taught to university students. In so doing, I ignore such subtleties as the existence of prominent individuals or schools of thought within a discipline that escape my generalizations. I justify this stance by reminding the reader that it is the core of the disciplines that must be changed, and the celebration of doctrinal diversity often serves to deflect attention away from the need for fundamental reform.

7. Dialogue with behavioral scientists has convinced me of the difficulty in maintaining a sustained scientific attitude when the BPC model is referred to as the "rational actor model." I will continue to use the latter term occasionally, although generally preferring the term "BPC model." Note that "beliefs" applies to both nonhuman species and human, as when we say, "We led the lion troop to believe there was a predator in the vicinity" or "We erected a mirror so that the fish believed it was being accompanied in predator inspection."

8. The fact that psychology does not integrate the behavioral sciences is quite compatible, of course, with the fact that what psychologists do is of great scientific value.

9. This argument was presented verbally by Darwin (1872) and is implicit in the standard notion of "survival of the fittest," but formal proof is recent (Grafen 1999; 2000; 2002). The case for frequency-dependent (non-additive genetic) fitness has yet to be formally demonstrated, but the informal arguments in this regard are no less strong.

10. For a more extensive analysis of the parallels between cultural and genetic evolution, see Mesoudi et al. (2006). I have borrowed heavily from that paper in this section.

11. I say "might" because in real life, individuals generally do not choose among lotteries by observing or contemplating probabilities and their associated payoffs, but by imitating the behavior of others who appear to be successful in their daily pursuits. In frequently repeated lotteries, the Law of Large Numbers ensures that the higher expected value lottery will increase in popularity by imitation without any calculation by participants.

12. I once presented this evidence to graduating seniors in economics and psychology at Columbia University, towards the end of a course that developed and used quite sophisticated probabilistic modeling. Many indicated in their essays that they did not believe the data.

13. In a careful review of the field, Shafir and LeBoeuf (2002) reject the performance error interpretation of these results, calling this a "trivialization" of the findings. They come to this conclusion by asserting that performance errors must be randomly distributed, whereas the errors found in the literature are systematic and reproducible. These authors, however, are mistaken in believing that performance errors must be random. Ignoring base rates in evaluating probabilities or finding risk in the Ellsberg two-urn problems are surely performance errors, but the errors are quite systematic. Similarly, folk intuitions concerning probability theory lead to highly reproducible results, although incorrect.

14. One could posit that the "type" of a player must include the player's probability distribution over the types of other players, but even such arcane assumptions do not solve the problem.

15. Current research is less sanguine concerning the important of reciprocal altruism in nonhumans (Hammerstein 2003).

Open Peer Commentary

Game theory can build higher mental processes from lower ones[1]

DOI: 10.1017/S0140525X07000593

George Ainslie

116A Veterans Affairs Medical Center, Coatesville, PA 19320.

george.ainslie@va.gov **www.picoeconomics.com**

Abstract: The question of reductionism is an obstacle to unification. Many behavioral scientists who study the more complex or higher mental functions avoid regarding them as selected by motivation. Game-theoretic models in which complex processes grow from the strategic interaction of elementary reward-seeking processes can overcome the mechanical feel of earlier reward-based models. Three examples are briefly described.

Gintis's call for unification is well reasoned, but some behavioral scientists may resist it because of a largely unspoken rift that

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

divides us into reductionist and anti-reductionist camps. The reductionists claim that people's various stated reasons for making choices – desire, duty, sympathy, ethics, and so on – ultimately depend on a unitary selective factor that operates in a single internal marketplace. The anti-reductionists do not have an alternative theory – pointedly – but shrink from the potential hubris of reductionist theories.

Reductionists infer the selective factor from the fact of choice itself (Premack 1959) and call it utility, satisfaction, reinforcement, reward, even "microwatts of inner glow" (Hollis 1983). Gintis follows the biologists in calling it fitness, or the expectation of fitness, but this usage confounds the selection of organisms – from which fitness is inferred – with the selection of behaviors within individuals (*proximate* as opposed to *ultimate* causality in his terms; see target article, sect. 1).[2] He is certainly a reductionist, but he does not say how the higher mental processes might be selected within individuals. For instance, his statements about internalized values being "constitutive," prevailing because of their moral value, and depending "in part on the contribution of values to fitness and well-being" (sect. 7) leave the role of the internal marketplace in their selection unclear.

Anti-reductionists have the same concern that may move the proponents of free will in philosophy, the fear that

> reductionism is a plague that grows proportionally as our society gets more sophisticated at controlling human behavior. We come to experience and conceptualize ourselves as powerless victims of mechanism, and thereby enter into a self-fulfilling prophecy. (Miller 2003, p. 63)

This fear is not entirely unfounded. For example, there is a lively debate about whether education in rational choice theory makes people less cooperative (Frank et al. 1996; Haucap & Just 2003). However, as Gintis points out, this education itself is probably erroneous. Likewise, the mechanical feel of reductionism may have come from some authors' procrustean application of simple experimental paradigms to complex human situations (e.g., Skinner 1948). Explicit hypotheses about how higher mental functions arise from lower ones might dispel robotic fantasies and clear the way for the unification Gintis envisions.

Elsewhere I have argued that rich human experience can be understood to arise from the interaction of simpler processes, without violence to its subtleties (Ainslie 2001; 2005). Hyperbolic discounting has the potential to motivate conflicting reward-based processes that can endure for long periods in a limited warfare relationship, giving an individual choice-maker many of the properties of a population of choice-makers. Just as "decision-making must be the central organizing principle of psychology" (target article, sect. 1.2.1), I submit that this limited warfare relationship among successively dominant interests in individuals must determine the basic nature of decision-making. The higher mental processes that are the starting point of cognitive psychology, sociology, and anthropology not only interact in ways that are clarified by game theory, as Gintis describes, but they also arise through game-theoretic mechanisms from simpler reward-seeking skills.

Three examples show the potential of this approach to go beyond the Skinner-box-writ-large: will, in the aspects of both strength (necessary for BPC's consistency; sect. 9.2) and freedom (necessary to meet antireductionist objections); vicarious reward, which interacts with will to motivate other-regarding preferences (sect. 10); and the construction of belief, for which Gintis seeks a mechanism in section 11 (see also sects. 6 and 7).

Will. Willpower can be understood as a person's[3] interpretation of her own choices in successive temptations as cooperations or defections in an intertemporal variant of repeated prisoner's dilemma (Ainslie 2001, pp. 78–104; 2005). Insofar as a person sees her choice about a current temptation as predicting how she will choose about similar future temptations, she adds the rewards for those choices to the rewards she can expect in the current choice – a perception that under hyperbolic but not exponential discounting gives her additional incentive to resist temptation. Given hyperbolic discounting, it is only by learning such perceptions that "the observed behavior of individuals with discount rates that decline with the delay" can "[become] choice consistent" (sect. 9.2). Thus, the will can be interpreted as the perception of a bargaining situation among a person's successive selves rather than as a faculty with inborn complexities. Furthermore, the sensitive dependence of repeated prisoner's dilemmas on individual choices makes their outcomes unpredictable from mere knowledge of their contingencies – even by the person herself – thereby arguably reconciling the experience of free will with determinism. This kind of bridge from the bottom upward in the hierarchy of complexity will not reduce the study of higher mental functions to something more molecular, but it can supply a context that connects them to basic motivational science.

Vicarious reward. Whatever way altruism and social virtues are selected by fitness, putatively their ultimate cause (sect. 10), Gintis and his cited authors address their proximate causes (rewards) only in terms of reciprocity. Hyperbolic discounting suggests how vicarious experience can be rewarding in its own right. The piece that has been missing in utility-maximizing theories of social utility is emotion. In contrast to conventional, conditioned reflex models of emotion, hyperbolic discounting permits emotion to be seen as a motivated process that taps endogenous sources of reward – transient reward alternating with inhibition of reward in the case of negative emotions, reward attenuated by anticipation and habituation in the case of positive emotions (Ainslie 2001, pp. 164–74, 179–86; 2005). Emotional reward does not physically require stimuli from the environment, but it still needs them in practice because it will habituate to the level of a daydream unless occasioned by environmental events that are both of limited frequency and partially unpredictable.

Various kinds of gambles, challenging tasks, and fictional stories are among the patterns that can meet these criteria, but the most apt should be the actual experience of other people. My hypothesis is that the experiences of other people acquire value in the internal marketplace of reward insofar as they are good occasions for emotion, and that both social virtues and social vices acquire value insofar as they support strategies of occasioning emotion, respectively in the long run and short run. The rewarding properties of the various emotions are undoubtedly shaped in evolution by their contribution to fitness. In the individual, however, emotion is a reward-producing behavior that produces more or less depending on how occasions pace its occurrence over time. Thus, in addition to self-regarding reciprocity, the stuff of sociology and anthropology is woven by emotion-cultivating processes that develop complex social skills to avoid habituation.

Construction of belief. Finally, Gintis says that "beliefs directly affect well-being" (sect. 11), by which he means that, apart from their instrumental value in getting other rewards, beliefs are rewarding in their own right. Social constructionists have long made this point, but have not said what constrains motivated belief; that is, what makes belief different from make-believe. Elsewhere I have argued that the noninstrumental value of beliefs is to occasion emotion (Ainslie 2001, pp. 175–79; 2005) and that the two kinds of value are often confounded because the limited occurrence of instrumental success also qualifies information predicting it as a good occasion for emotion (Lea & Webley 2006; Ainslie 2006). "Transcendental beliefs" (sect. 6) are a large category of emotionally useful belief that is made unique for the individual not by instrumental accuracy but by cultural consensus. Such beliefs have to be transmitted in "conformist" fashion lest they lose their uniqueness and thereby weaken their value as occasions for emotion – but they still survive only insofar as they produce individual reward. Likewise, although a person is apt to shed

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

suggested norms that are not useful to her as boundaries against temptation (criteria for cooperation in her intertemporal prisoner's dilemmas; see my subsection *Will* above), she will find that the ones she believes to be uniquely dictated by fact ("internalizes" – sect. 7) are the most effective, as Gintis observes.

Just as societies are constructed by individuals interacting strategically, so too these individuals are constructed by basic reward-seeking processes that also interact strategically. However, maximizing reward implies neither selfishness nor determination by external contingencies.

NOTES

1. The author of this commentary is employed by a government agency and as such this commentary is considered a work of the U.S. government and not subject to copyright within the United States.

2. Of course, the factor that selects behaviors within an individual must in turn have been selected in the species by its effect on fitness; but it may still lead her well astray from fitness, as witness cocaine and birth control.

3. It is possible, but doubtful, that some nonhuman animals have sufficient theory of mind to use their own current choices as predictive cues.

The behavioral sciences are historical sciences of emergent complexity

DOI: 10.1017/S0140525X0700060X

Larry Arnhart

Department of Political Science, Northern Illinois University, DeKalb, IL 60115.
larnhart@niu.edu **Darwinianconservatism.blogspot.com**

Abstract: Unlike physics and chemistry, the behavioral sciences are historical sciences that explain the fuzzy complexity of social life through historical narratives. Unifying the behavioral sciences through evolutionary game theory would require a nested hierarchy of three kinds of historical narratives: natural history, cultural history, and biographical history.

Evolutionary biology and the behavioral sciences are historical sciences of emergent complexity. By contrast, physical sciences such as physics and chemistry are nonhistorical sciences of reductive simplicity (Mayr 1996; 2004; Morowitz 2002). And yet many social scientists – particularly economists – have taken physics as the model for all science, and they have tried to unify the behavioral sciences as founded on social physics (Mirowski 1989). Gintis correctly rejects this approach as he tries to unify the behavioral sciences as founded on evolutionary history.

Pursuing social physics sacrifices accuracy for the sake of precision, because it ignores the fuzzy complexity of social reality. Pursuing evolutionary history sacrifices precision for the sake of accuracy, because it recognizes that fuzzy complexity (Blalock 1984). Gintis's proposal rightly rejects the first in favor of the second. But in doing so, he does not go far enough in explicitly recognizing the fuzzy complexity in the science that he proposes. He should stress more than he does that the behavioral sciences are historical sciences of emergent complexity that move through a nested hierarchy of three kinds of historical narratives: natural history, cultural history, and biographical history.

I can illustrate what I mean through the topic of war. Charles Darwin believed that warfare was crucial for the evolution of human social and moral capacities (Darwin 1871). Similarly, Gintis suggests that the evolution of strong reciprocity could have depended on lethal combat, so that groups with high levels of strong reciprocity would have been more likely to prevail in war against their opponents (Gintis et al. 2005b). Considering the importance of war, we might ask how the behavioral sciences as rooted in evolutionary game theory should explain the origins of war in general and of wars in particular circumstances (such as, for example, the American invasion of Iraq in the spring of 2003).

Natural history. Except for historical disciplines such as cosmology and geology, physical scientists study physical phenomena without reference to their history. But behavioral scientists cannot explain the behavioral phenomena they study without historical narratives; and if they adopt Gintis's proposal, they will have to start with the natural evolutionary history of the human species. The very possibility of a science of human behavior assumes enough stability in human nature so that the scientist can explain behavior as manifesting probabilistic propensities. Gintis would say that those behavioral propensities ultimately arose from human evolutionary history.

So, if strong reciprocity is a propensity of the human species, we should be able to explain it as a product of Darwinian evolution. But, unlike theories in physics and chemistry, evolutionary theory cannot be tested directly by laboratory experimentation, because we cannot replicate the history of evolution in the laboratory. Instead, we must formulate historical narratives of evolutionary history, and then test those narratives by seeing how far they conform to the relevant evidence and logic. Gintis suggests various scenarios by which evolutionary group selection would favor strong reciprocity as enhancing fitness (Gintis et al. 2005b).

If warfare is important to these scenarios, then we would have to decide whether the pattern of warfare in human evolutionary history supports the historical narrative. In fact, some social scientists have argued that there has been a history of warfare in human evolution that would confirm such a narrative (Rosen 2005; Thayer 2004).

But, unlike the deterministic laws of the nonhistorical sciences, such historical narratives would lead us only to probabilistic regularities. For example, we might conclude that since evolutionary history has favored a propensity for lethal fighting among males more than among females, the propensity for war will be stronger among men than among women. But such a propensity will be highly variable and contingent on circumstances.

In contrast to the motivational reductionism of *Homo economicus*, Gintis sees human nature as both self-regarding and other-regarding. But, like Adam Smith (1759/1982), I would argue that we need to recognize even more motivational complexity to account for the hundreds of human psychological universals clustered around 20 natural desires (Arnhart 1998; 2005; Brown 1991; Westermarck 1906).

These evolved natural desires constrain but do not determine cultural learning and individual judgments. This seems to be what Gintis has in mind when he says that sociology "systematically ignores the limits to socialization" (sect. 7), because it assumes a "blank slate" that denies human nature.

Cultural history. Gintis argues that framing effects are unavoidable in game theory experiments, because when subjects enter an experiment, they necessarily apply the social and moral concerns of "everyday life" in the culture in which they live. Strong reciprocity and other natural propensities will vary across cultures, and subjects in different cultures will differ in their experimental game play because of differences in the cultural history of their social lives (Gintis et al. 2005b).

To explain why George W. Bush decided to invade Iraq in the spring of 2003, behavioral scientists would have to consider how the cultural institution of the presidency gave him the power to make war in the circumstances that he faced. It might be a natural propensity of human beings to live in societies with dominance hierarchies and to defer to dominant leaders, particularly in war. The willingness of citizens to risk their lives in war is a heroic manifestation of the human disposition to other-regarding behavior. But that natural propensity in the United States is channeled through the constitutional and social history of the American president as commander-in-chief.

Biographical history. Gintis cites Barrington Moore (1978) as showing how the natural desire for justice as reciprocity motivates moral and political reform. Moore indicates, however, that social reform movements depend on the

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

decisions of individual leaders, and therefore social historians must study the actions of those individuals to decide their moral and political responsibility for historical events. Within the constraints set by natural propensities and cultural circumstances, the judgments of unique individuals in positions of responsibility can decisively determine the course of human history. This makes human behavioral history more complex and unpredictable than anything studied by the physicist or chemist.

To explain fully why President Bush launched the American invasion of Iraq, the behavioral scientist would have to consider not only the natural history of war in human evolution and the cultural history of presidential war in the United States, but also the biographical history of President Bush and those who influenced his decision.

Social complexity in behavioral models

DOI: 10.1017/S0140525X07000611

R. Alexander Bentley

Department of Anthropology, Durham University, Durham DH1 3HN, United Kingdom.

r.a.bentley@durham.ac.uk

http://www.dur.ac.uk/anthropology/staff&/profiles/?id=2570

Abstract: Although the beliefs, preferences, and constraints (BPC) model may account for individuals independently making simple decisions, it becomes less useful the more complex the social setting and the decisions themselves become. Perhaps rather than seek to unify their field under one model, behavioral scientists could explore when and why the BPC model generally applies versus fails to apply as a null hypothesis.

There could be no better motivation for a *BBS* target article than to unify the behavioral sciences; but with the myriad competing perspectives, this objective may be fundamentally unattainable. Gintis, for instance, mainly advocates a classical economic and game-theoretic view of the topic, in which there are several debatable assumptions and neglected alternative approaches, involving the following:

Cultural variability and relativity of social costs and benefits. In the beliefs, preferences, and constraints (BPC) model, the assumption that human decisions have an optimal value (in the sense of a fitness-enhancing payoff) neglects how many behaviors are highly culturally dependent and individually variable, even within small-scale societies living in similar environments (e.g., Cronk 1999; Henrich et al. 2006). So, although the basic equations of the BPC model are internally consistent, to what behaviors do they apply? Recent game-theoretical field experiments (Henrich et al. 2006) in non-Western societies compellingly show a variety of different culturally dependent solutions to social dilemmas.

Complexity. The assumption in the BPC model – that payoffs are predictable from one event to the next – may potentially be true; for example, of hunting and gathering in a consistent environment. But most social benefits depend on what other actors are doing, and by changing from one event to the next, these benefits can easily become analytically intractable. Also, rather than a discrete choice between doing one thing or the other (e.g., defect or cooperate), most real-world choices involve many options (e.g., what friends to keep, what job to pursue) and often are not even discrete (e.g., where on the landscape to hunt or to fish). Complex choices can be fundamentally different from simple two-choice scenarios, when the problem becomes literally unpredictable from an individual-centered analysis. In physics, for example, the two-body orbit problem is analytically predictable, but the three-body problem is not. How much, then, do two-choice models tell us about the behavior of many people, each of whom is repeatedly deciding from multiple (potentially very many) choices?

BPC would seem to work best in cases where the complexity of choices is lowest (Winterhalder & Smith 2000). When it becomes impossible to estimate a payoff probability distribution for each individual at each successive event, such problems are then better addressed by computer simulation (e.g., Axelrod 1997a; Conte et al. 1997; Gilbert & Troitzsch 1999; Lansing 2006; see also *Journal of Social Sciences and Simulation*). Also relevant is the general field of "econophysics," engagingly introduced by Ormerod (1998; 2005), and other general overviews of non-equillibrium dynamics in collective behavior (e.g., Ball 2004).

Random copying versus conformity. Gintis discusses imitative behavior, but mostly in terms of conformism or prestige-biased imitation (cf. Boyd & Richerson 1985; Henrich & Boyd 2001; Henrich & Gil-White 2001). As demonstrated by empirical data from modern and prehistoric societies (e.g., Bentley & Shennan 2003; Neiman 1995; Salganik et al. 2006; Shennan & Wilkinson 2001; Simkin & Roychowdhury 2003), other forms of imitation include copying within the structured constraints of a social network (e.g., Granovetter 2003; Newman et al. 2006; Pool & Kochen 1978; Wasserman & Faust 1994) and copying other individuals at random, akin to the random drift model in population genetics (e.g., Bentley & Shennan 2005; Eerkens & Lipo 2005; Hahn & Bentley 2003; Herzog et al. 2004; Lipo et al. 1997).

Random copying and conformity are quite different, and can have significantly different effects. Copying a randomly selected individual is not the same as making an informed decision about the most common behavior (or prestigious individual) to imitate. Whereas random copying leads to a power law distribution in the popularity of choices, with the most popular choice arising simply by chance (Bentley et al. 2004; Hahn & Bentley 2003; Simkin & Roychowdhury 2003), conformist or prestige bias would more likely give rise to "winner-take-all" distribution, where there can be at least some explanation for the predominant choice (Bentley & Shennan 2003; Watts 2002).

The BPC model may be an effective null hypothesis for the behavior of individuals making independent, "either-or" decisions. The more that people's decisions depend on what others are doing, however, and the more choices they have, the weaker the BPC model becomes, and the less it serves as a unifying principle. Exploring the societal transitions between these realms – from where BPC holds to where it does not – could be truly fascinating.

Towards uniting the behavioral sciences with a gene-centered approach to altruism

DOI: 10.1017/S0140525X07000623

R. Michael Brown[a] and Stephanie L. Brown[b]

[a]*Department of Psychology, Pacific Lutheran University, Tacoma, WA 98447;* [b]*Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109.*

brownrm@plu.edu **stebrown@med.umich.edu**

Abstract: We support the ambitious goal of unification within the behavioral sciences. We suggest that Darwinian evolution by means of natural selection can provide the integrative glue for this purpose, and we review our own work on selective investment theory (SIT), which is an example of how other-regarding preferences can be accommodated by a gene-centered account of evolution.

We wholeheartedly support the ambitious goal of unification within the behavioral sciences. Towards this end, we agree that Darwin's theory of evolution holds great promise as an organizing and integrating framework and may have the potential to be as

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

generative for the behavioral sciences as it has been for the biological sciences. We also applaud Gintis for suggesting that other-regarding preferences may be a key feature of unification, that decision-making should be the central concern of cognitive psychology, and that game theory has great potential for modeling *human* as well as animal behavior.

Unfortunately, Gintis does not include in his proposal a recipe for unification, other than to say, "We must recognize evolutionary theory (covering both genetic and cultural evolution) as the integrating principle of behavioral science" and that the rational actor model must be "broadened to encompass other-regarding preferences." (sect. 12, para. 3). Nor does Gintis acknowledge the daunting challenges to unification, other than to provide a list of "misconceptions" concerning the rational actor model and game theory, objections that he summarily dismisses in laundry list fashion. But there are, in fact, real barriers to the Gintis agenda for unification, across the behavioral sciences and within each of them. In psychology, for example, evolutionary theory continues to be a hotly debated and contested perspective that engenders as much controversy as promise. To wit, academic troops still battle over whether evolution can inform our understanding of gender differences (e.g., Eagly & Wood 2003), social bonds (Berscheid 2006; R. Brown & S. Brown 2006), or altruism (Batson 2006). And the controversy does not stop there. The mere mention of altruistic motivation can engender ridicule (Batson 1997; Neuberg et al. 1997), perhaps owing to the surprisingly *unified* (across discipline), fortified, and long-standing belief that *all* human and animal behavior is motivated by self-interest (psychological hedonism).

Even those eager to jump on the evolution bandwagon and embrace an other-regarding perspective have questions – *Which* bandwagon? *Which* perspective? Gintis only hints at answers, but elsewhere (e.g., Gintis 2000; Gintis et al. 2003), he and his colleagues have made it clear that they subscribe to views of evolution and other-regarding preferences that are themselves steeped in controversy. Their arguments for "true altruism," in which helpers sacrifice *inclusive* fitness for the good of the group, rest heavily on the controversial notion of group selection, an assumption that is decidedly out of the mainstream of evolutionary biology (Alcock 2001). And the case they cite as evidence for ultimate altruism (and group selection) – strong reciprocity – is neither decisive nor compelling proof of either proposition (Burnham & Johnson 2005; Hagen & Hammerstein 2006; Sanchez & Cuesta 2005).

If evolution is to unify the behavioral sciences, then it may be important, at least initially, to settle on the version of evolution that has, in fact, served as integrative glue for the *biological* sciences: Darwinian evolution by means of natural selection, informed and modified by discoveries in genetics and by the insight that the fundamental target of selection is the gene, not the group, the species, or even the individual. This gene-centered view of evolution is accepted by the overwhelming majority of evolutionary biologists and by scientists in other disciplines who study the evolution of behavior. And, contrary to what some behavioral scientists might think, the gene-centered view of evolution can and does support other-regarding preferences; there is no need to buy into the less parsimonious and more controversial notion of group selection.

Our own contribution in this area – *selective investment theory* (SIT) (S. Brown & R. Brown 2006) – provides an illustrative example of how other-regarding preferences can be accommodated by a gene-centered account of evolution. SIT was formulated to help explain, from a gene-centered evolutionary perspective, the motivational basis for high-cost altruism (e.g., parental care, defense of family members as well as genetically unrelated coalition partners). How is this kind of giving/helping accomplished – especially in view of conflicting self-centered motives that are evolutionarily ancient, were vital to our emergence as a species, and continue to drive our behavior today? SIT holds that it is the *social bond* – the glue of close interpersonal relationships – that evolved to discount the risks of engaging in high-cost altruism. More specifically, SIT views the social bond as an emotion-regulating mechanism that functions to override self-interest and facilitate costly investment in others.

A key component of SIT is that (a) if social bonds evolved to motivate high-cost altruism, then (b) the formation of social bonds must have occurred only between individuals who were dependent on one another for reproductive success, a condition we call *fitness interdependence*. Relationships in which individuals are dependent on one another for survival and reproduction provide givers with a "genetic safety net," making them resistant to exploitation. As de Waal (1996, p. 27) notes: "There can be little doubt that in many species the strong can annihilate the weak. In a world of mutual dependency, however, such a move would not be very smart." The "reproductive insurance" provided by fitness interdependence makes it a logically appealing prerequisite for forming social bonds, which function to facilitate high-cost altruism. Evidence from game theory confirms a link between fitness interdependence and the evolution of altruistic behavior. And there are considerable data, from both nonhuman and human species, that are consistent with the central tenets of SIT.

SIT is based on the assumption that altruism was necessary for ensuring the survival, growth, and reproduction of interdependent ancestral humans. Hence, the spread of altruism is no mystery from a gene-centered perspective, even altruism directed to genetically unrelated humans. Selfish genes *can* produce other-regarding preferences. In view of this, the evolutionary model of choice for unifying the behavioral sciences should be obvious. It is the same model that has organized and catalyzed discoveries in the biological sciences for 40 years.

Evolutionary theory and the social sciences

DOI: 10.1017/S0140525X07000635

Robert L. Burgess and Peter C. M. Molenaar

Human Development and Family Studies, Pennsylvania State University, University Park, PA 16802.

rlb8@psu.edu **pxm21@psu.edu**

Abstract: Gintis's article is an example of growing awareness by social scientists of the significance of evolutionary theory for understanding human nature. Although we share its main point of view, we comment on some disagreements related to levels of behavioral analysis, the explanation of social cooperation, and the ubiquity of inter-individual differences in human decision-making.

Gintis's basic thesis is that the principles of evolutionary theory have the capacity to integrate the various behavioral and social sciences. We are in full agreement with this thesis. Indeed, because of the complexity of human behavior and its development, it is essential that research be theoretically grounded and that care be taken to integrate biological as well as environmental factors in our explanations of this complex topic. Evolutionary theory is singularly well-placed to accomplish this task. Why? Because it is the most general theory we have in the life sciences and, therefore, has the greatest potential to unify these various disciplines. We also agree that recognizing evolutionary theory as the most general theory in the life sciences is not to deny the significance of the allied disciplines of anthropology, economics, history, psychology, or sociology, nor their "middle-range" theories. The central intellectual problem of those fields is not *analytic*, that is, discovering new and fundamental general theories. Rather, their problems are *synthetic*: showing how genes and environments in accordance with evolutionary

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

principles combine to produce our common human nature and the diversity of ways in which that nature is manifested (Burgess 2005).

Differences do exist, however, about how evolutionary theory can best be used to explain human behavior and its development in different contexts, and how it can integrate the various behavioral and social sciences. For example, Barkow et al. (1992) have maintained that human behavior is influenced by evolved *domain-specific* mechanisms, rather than by *domain-general* mechanisms that generalize across multiple behavioral domains. Domain-specific mechanisms are said to have achieved their exalted status because they solved recurrent adaptive problems faced by ancient humans throughout history. Because those problems were many and diverse, our minds are equipped with a variety of domain-specific psychological mechanisms. The assumption is that the human mind could not possibly be composed of all-purpose domain-general mechanisms. That, these authors imply, was the folly of operant conditioning in psychology.

A different perspective is taken by evolutionary anthropologists (e.g., Flinn 2005; Irons 1979), who emphasize the role played by variable ecological factors in influencing adaptive behavior. Similarly, the "dual inheritance" approach views cultures and genes as providing separate but linked systems affecting evolutionary change (e.g., Boyd & Richerson 1985). Finally, evolutionists interested in developmental questions (e.g., Alexander 1979; various contributors to Burgess & MacDonald 2005) emphasize the importance of domain-general as well as domain-specific psychological mechanisms. This position is based on doubt that there ever was a single environment that was common to all ancient humans. From this perspective, the very uncertainty and diversity of the environments that our ancestors faced led to the selection of psychological mechanisms of sufficient generality to permit adaptation to changing environments.

Apart from our basic agreement with Gintis, there are several points with which we must respectfully disagree. First, he expresses his objection to reductionism; yet theory construction in science, wherein one attempts to explain complex phenomena by deriving them from more general principles, is intrinsically a reductionist process. That it is so does not mean replacing one field of knowledge with another, but rather, linking them. In biology, explanation is generally felt to occur on four complementary levels of analysis, and these different levels reflect the fact that the various behavioral disciplines are divided less by the theories they employ than by the problems they address. These four levels are the evolutionary history of a trait, its adaptive function, the development of the trait in an individual's life span, and the specific and proximate mechanisms that cause a trait to be expressed (Tinbergen 1963). A common thread runs through each of these levels: evolutionary history and adaptiveness being more general than development and proximate antecedents. Genetic processes are involved at each level. Developmental and proximate mechanisms can be deduced from (reduced to) the first two levels under empirically specified "given conditions."

Second, Gintis argues that it is highly implausible that cooperation is a product of individuals pursuing their own personal goals. We agree that pan-human traits evolved in a social context. Nevertheless, these very traits are important precisely because they are experienced by "self-regarding rational agents" and individual actions are influenced accordingly. And, we humans typically find ourselves in situations where we are dependent upon the actions of others in order to attain our own individual goals. It is this state of mutual dependence that leads to the "norm of reciprocity." Relationships among kin are unique, and, indeed, rules of morality probably evolved therein. The link between moral behavior and kin relations is seen in the ancient Arab proverb: "I against my brother; my brother and I against our cousin; my brother, my cousins, and I against the world." Beyond kin-based altruism, moral behavior is often sustained through reciprocal altruism and by coercion in more complex societies (van den Berghe 1990).

Our third point concerns the extreme individual variability observed in behavioral studies of decision-making and choice. Luce (2000, p. 29) points out the occurrence of substantial individual differences necessitating that "Each axiom should be tested in as much isolation as possible, and it should be done in-so-far as possible for each person individually, not using group aggregations." This is a common finding; a recent empirical and theoretical analysis of binary choice behavior reports the presence of "extremely large individual differences" (Erev & Barron 2005, p. 925).

These large individual differences, and the consequent necessity to use person-specific (time series) designs and data analysis techniques, raise the important issue concerning the relationship between inter-individual variation (i.e., the type of between-units variation that is supposed to underlie evolution) and intra-individual variation (i.e., the type of within-unit time series variation that underlies individual learning and development). It turns out that, in general, the structure of inter-individual variation of some phenotype (as assessed, for example, in standard structural equation modeling) is unrelated to the structure of intra-individual variation of the same phenotype (as assessed, for example, in multivariate time series analysis). Only if the phenotypic process under consideration is ergodic – that is, has invariant statistical characteristics across subjects and time – are the structures of inter- and intra-individual variation asymptotically the same (cf. Molenaar 2004). The criteria for ergodicity imply that all processes with time-varying statistical characteristics, such as learning and development, are non-ergodic – that is, their inter-individual structure of variation is unrelated to the intra-individual structures of variation. This consequence of the so-called classical ergodic theorems has major implications for psychometrics (cf. Molenaar 2004; see also Borsboom & Dolan 2006) and, along with the other issues we raised, needs to be addressed explicitly in the context of Gintis's thesis.

There is nothing too surprising here. The concept of the phenotype, as a product of genotypes, acknowledges the flexible and variable ways in which individuals respond to differing environmental circumstances and developmental experiences. The ability to adapt to different environments and to learn different things is a product of natural selection; hence, learning, development, and phenotypes depend upon evolutionary history and principles. To fully understand any behavioral phenomenon, we need to address all four of Tinbergen's complementary levels of analysis.

Against the unification of the behavioral sciences

DOI: 10.1017/S0140525X07000647

Steve Clarke

Centre for Applied Philosophy and Public Ethics, Charles Sturt University, Canberra, ACT 2601, Australia; and Program on the Ethics of the New Biosciences, James Martin 21st Century School, University of Oxford, Oxford, OX1 1PT, United Kingdom.

stephen.clarke@anu.edu.au

http://www.cappe.edu. au/people/clarst/clarst.htm

Abstract: The contemporary behavioral sciences are disunified and could not easily become unified, as they operate with incompatible explanatory models. According to Gintis, tolerance of this situation is "scandalous" (sect. 12). I defend the ordinary behavioral scientist's lack of commitment to a unifying explanatory model and identify several reasons why the behavioral sciences should remain disunified for the foreseeable future.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Gintis aims to unify the currently very disunified behavioral sciences, advocating the general adoption of his "beliefs, preferences, and constraints (BPC) model" of human behavior. The BPC model is a variant of the "rational actor model," ubiquitous in economics (sect. 1.2.2). According to Gintis, it is "scandalous" that the different behavioral sciences currently offer up "partial, conflicting, and incompatible models" and have done so for most of the twentieth century (sect. 12, para. 1). Here I defend the ordinary behavioral scientist's lack of commitment to any one unificatory model, identifying several reasons why the behavioral sciences are better off remaining disunified for the foreseeable future.

According to Gintis, the last serious proposal for the unification of the behavioral sciences was presented in 1951 (Note 2). This claim suggests a narrow construal of what counts as a serious proposal for the unification of the behavioral sciences. The structuralist social theories developed by Althusser, Poulantzas, and others, in the 1960s and 1970s, can be understood as attempts to unify the behavioral sciences (Resch 1992). While the rational actor model locates the agent at the center of social explanation, structuralists downplay the importance of agency and instead emphasize the importance of a socially determined unconscious in explaining individual behavior. Attempts to reconcile social structure with agency, such as those due to Bourdieu (Harker et al. 1990), the later Sartre (Levy 2002, pp. 119–44), and the critical realist Bhaskar (1979), can also be understood as attempts to unify the behavioral sciences.

There are several models and proto-models for the unification of the behavioral sciences currently available. Why think that behavioral scientists, most of whom appear happy to do without any particular unificatory model, would be better off accepting one of these? Apart from alluding to benefits that follow from breaking down disciplinary boundaries (sect. 12, para. 2), Gintis does not address this question. It seems plausible to think that Gintis simply assumes that unificatory power and explanatory strength go hand in hand. And indeed there is a long tradition of relating the two (Kitcher 1989).

The project of unifying the sciences was pursued by many in the middle third of the twentieth century. However, it has fallen out of favor, at least in philosophy, mostly as a result of unanswered criticisms of the various proposals to unify particular sciences (Wylie 1999). A far-reaching criticism of unificatory models of explanation in the natural sciences comes from Cartwright (1999), who argues that the apparent success of simple explanatory models in the natural sciences results from these being heavily idealized and distantly abstracted from the complexity of reality. Cartwright also asks us to contemplate the possibility that nature is at bottom "dappled" and that there may be no descriptively accurate unified model of reality to be had. Prominent advocates of explanatory unification, such as Kitcher, now accept that reality may be intrinsically disunified in at least some of its aspects and advise us to accept unificatory explanations only when and where these remain descriptively accurate (Kitcher 1999, p. 339). Social reality is at least as complicated as physical reality, and it seems plausible to think that simple ideal models that may be used to explain social reality, such as the BPC model, have the explanatory reach that they have only because they are abstracted away from the messiness of reality. If social reality is disunified, then explanatory unification in the behavioral sciences can only be had at the cost of descriptive inaccuracy.

But even if social reality is unified, it may still be a bad idea for contemporary behavioral scientists to collectively adopt one unifying explanatory model. The adoption of a particular model poses three problems. First, because the different behavioral sciences have developed in incompatible ways, the unification of the behavioral sciences would involve the abandonment of much work that does not fit easily into the unifying framework adopted. Radin (1996) argues that the expansion of the rational actor model of explanation into areas of behavioral science in which it has not traditionally been employed would cause a significant loss of "local knowledge." Crucially, she argues that the rational actor model has no capacity to account for incommensurable values. Because Gintis's BPC model is a variant of the rational actor model, Radin's criticisms apply straightforwardly to it.

Second, the general acceptance of a particular unifying model may prevent new perspectives from being developed, from which criticisms of the presuppositions of the accepted model might be made. Gintis devotes much space to showing how the BPC model can be reconciled with evidence of apparent failures of people to behave rationally that has been identified by Tversky and Kahneman and others. But a more serious concern is whether research that challenges the presupposition that people generally act rationally would have been conducted in a unified behavioral science in which the BPC model was adopted. A unified behavioral science could be expected to have many of the characteristics of a Kuhnian paradigm, as we find in the natural sciences. This would bring some benefits to the behavioral sciences. However, it would also involve a serious disadvantage. As Kuhn (1970, pp. 35–42) argues, under normal conditions, in a unified discipline researchers are severely discouraged from attempting to conduct work that threatens to undermine accepted background assumptions.

Finally, the general acceptance of one unifying model of the behavioral sciences would presumably involve the cessation of work intended to advance the case for other unifying models of the behavioral sciences. There are a multiplicity of unifying models and proto-models of the behavioral sciences available, none of which has won anything close to general acceptance. Plausibly, this is because none of these offers explanations that are clearly better than those offered by its rivals. Given this state of affairs, it would be extremely reckless for behavioral scientists as a whole to conduct work only within the framework of one such model. To do so would involve abandoning work within other frameworks that might enable superior explanations of behavior to be developed in the future.

ACKNOWLEDGMENTS

Thanks to Neil Levy, Kate MacDonald, Terry MacDonald, and Seumas Miller for helpful comments.

Love is not enough: Other-regarding preferences cannot explain payoff dominance in game theory

DOI: 10.1017/S0140525X07000659

Andrew M. Colman

School of Psychology, University of Leicester, Leicester LE1 7RH, United Kingdom.

amc@le.ac.uk

http://www.le.ac.uk/home/amc

Abstract: Even if game theory is broadened to encompass other-regarding preferences, it cannot adequately model all aspects of interactive decision making. Payoff dominance is an example of a phenomenon that can be adequately modeled only by departing radically from standard assumptions of decision theory and game theory – either the unit of agency or the nature of rationality.

Gintis rests his attempt to unify the behavioral sciences on a claim that "if decision theory and game theory are broadened to encompass other-regarding preferences, they become capable of modeling all aspects of decision making" (Abstract). This claim seems unsustainable in relation to many aspects of both individual and interactive decision making, but I shall confine my comments to just one, namely the payoff-dominance phenomenon. The simplest illustration of it is the Hi-Lo matching game depicted in Figure 1.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

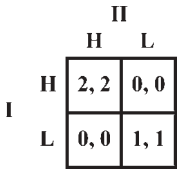


Figure 1 (Colman). Payoff matrix of the Hi-Lo game.

Player I chooses between rows H and L, and Player II independently chooses between columns H and L. The pair of numbers in the cell where the chosen row and column intersect are the payoffs to Player I and Player II, respectively. The Hi-Lo game is a pure coordination game, because the players' interests coincide exactly and they are motivated to match each other's strategy choices. This payoff structure might apply to an incident in a football game, for example, when Player I can pass the ball either left or right for Player II to shoot for goal, and Player II can move either left or right to intercept it. If the chances of scoring are better if both choose left than if both choose right, and zero if they make non-matching choices, then their problem can be modeled as a Hi-Lo game (Bacharach 2006, pp. 124–27; Sugden 2005). Many other dyadic interactions have this simple strategic structure, and payoff dominance is also a property of more complicated games.

In game theory, payoffs represent utilities, but for the purposes of the argument that follows, we may interpret them simply as monetary payoffs – dollars, let us say. A fundamental assumption of orthodox game theory is that players are rational, in the sense of invariably acting to maximize their own (individual) expected payoffs, relative to their knowledge and beliefs at the time. This merely formalizes the notion that decision makers try to do the best for themselves in any circumstances that arise.

In the Hi-Lo game, it is obvious that rational players should choose H, and experimental evidence confirms that that is what (almost) everyone does in practice (Gold & Sugden, in press; Mehta et al. 1994). The HH outcome is in Nash equilibrium, because each player's strategy is a best reply to the co-player's; and this equilibrium is payoff dominant, in the sense that it yields both players a strictly higher payoff than the LL equilibrium, where strategies are also best replies to each other. Nevertheless, it is strange but true that game theory provides no justification for choosing H (Bacharach 2006, Ch. 1; Casajus 2001; Colman 2003a; Cooper et al. 1990; Crawford & Haller 1990; Harsanyi & Selten 1988; Hollis 1998; Janssen 2001). A player has no reason to choose H in the absence of a reason to expect the co-player to choose H, but the symmetry of the game means that the co-player faces the same dilemma, having no reason to choose H without a reason to expect the co-player to choose it. This generates an infinite regress that spirals endlessly through loops of "I expect my co-player to expect me to expect. . ." without providing either player with any rational justification for choosing H.

Other-regarding preferences provide no help in solving this problem, notwithstanding Gintis's claim. The usual way of modeling other-regarding preferences, although Gintis does not spell this out, is by transforming the payoffs of any player who is influenced by a co-player's payoffs, using a weighted linear function of the player's and the co-player's payoffs. This technique was introduced by Edgeworth (1881/1967, pp. 101–102) and has been adopted by more recent researchers, such as Rabin (1993) and Van Lange (1999). It alters the strategic structure of the well-known Prisoner's Dilemma game radically, providing a reason for cooperating where there was none before, but it leaves the Hi-Lo game totally unchanged. For example, suppose that both players attach equal weight to their own and their co-player's payoffs, then Player I's payoff for joint H choices is transformed from 2 to $(2 + 2)/2 = 2$, but this is exactly the same as before. The transformed, other-regarding payoff is identical to the untransformed, self-regarding payoff; and the same applies to all other payoffs of the game. This game is unchanged by other-regarding payoff transformation, and other-regarding preferences cannot solve the payoff-dominance problem in other games.

This is just one illustration of the fact that game theory cannot model all aspects of strategic decision making, even if it is broadened to encompass other-regarding preferences. The payoff-dominance phenomenon, illustrated by the Hi-Lo game, cannot be modeled within the framework of orthodox game theory (Colman 2003a; 2003b). The only valid solutions, as far as I am aware, involve either abandoning the assumption of individual agency that is fundamental to both decision theory and game theory (Bacharach 1999; 2006; Sugden 1993b; 2005) or assuming that players use a form of evidential reasoning that violates orthodox assumptions of rational decision making (Colman & Bacharach 1997; Colman & Stirk 1998).

It is worth commenting that any evolutionary game-theoretic model that operates by adaptive learning in a non-rational process of mindless trial and error would tend to converge on the payoff-dominant equilibrium in a game such as Hi-Lo, although this cannot explain why human players choose it in a one-shot game. But the version of evolutionary game theory favored by Gintis incorporates a rational actor "BPC" model in which the brain, as a decision-making organ, follows the standard principles of rationality. Gintis believes this to be a basic insight that is surprisingly "missing from psychology," and he devotes the whole of section 9 of his target article to defending it against its critics.

I must comment, finally, on Gintis's surprising assertion that the Parsons-Shils general theory of action was "the last serious attempt at developing an analytical framework for the unification of the behavioral sciences" (Note 2). There have been other attempts, of which the theory of operant conditioning (Ferster & Skinner 1957) is surely the most comprehensive, successful, and enduring (Dragoi & Staddon 1999), and it even underpins the Pavlov strategy of evolutionary games (Nowak & Sigmund 1993).

ACKNOWLEDGMENT

The preparation of this commentary was supported, in part, by an Auber Bequest Award from the Royal Society of Edinburgh, Scotland.

The place of ethics in a unified behavioral science

DOI: 10.1017/S0140525X07000660

Peter Danielson

W. Maurice Young Centre for Applied Ethics, University of British Columbia, Vancouver, BC V6T 1Z2, Canada.

pad@ethics. ubc.ca **http://www.ethics. ubc.ca/people/danielson/**

Abstract: Behavioral science, unified in the way Gintis proposes, should affect ethics, which also finds itself in "disarray," in three ways. First, it raises the standards. Second, it removes the easy targets of economic and sociobiological selfishness. Third, it provides methods, in particular the close coupling of theory and experiments, to construct a better ethics.

The target article proposes to unify behavioral science around evolutionary game theory. Although Gintis makes no explicit reference to ethics (except, perhaps, as part of philosophy), it is clear that concerns central to ethics – accounting for and, we hope, justifying prosocial attitudes – are also central to his proposal. On Gintis's account, *unified* behavioral science (UBS) is quite friendly to ethics. It is centered on choice, gives a central role to the normative ideal of rationality, and makes a case for moralized preferences as a product of evolution. Here I argue that a unified behavioral science should lead to, if not include, a unified science of ethics. In particular, I expect the change

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

promoted by Gintis to have three beneficial effects on the field of ethics.

1. Raising standards. Facing non-unified social science in "abiding disarray" allowed naturalistically inclined ethicists to pick and choose between frameworks. Some theorists preferred rational choice for its normative focus, others favored "qualitative" social science for its sensitivity to context and, often, its moral tone. Still others were discouraged by strong historical differences between the approaches in social science disciplines. In effect, non-unified social science lowers the standards that those working in the field, including ethicists, need to meet (Abel 2003). One consequence of unification is that it may be more evident that ethics, like the behavioral sciences, needs to articulate and defend "a model of individual human behavior" and (Abel adds) a model of social interaction.

Put another way, like social science as portrayed by Gintis, the field of ethics is also in "disarray." Well-regarded work ranges from nearly axiomatic exposition of various normative principles with disdain for empirical evidence – Kagan (1991) and Gauthier (1986) stand out at this extreme – to reliance on very concrete qualitative data, at the other extreme: witness the title of Hoffmaster (1993): "Can Ethnography Save the Life of Medical Ethics?" As I discuss in my subsection 3, UBS bodes ill for both of these methodological extremes, by demonstrating the power of tightly coupling theories formulated as models with experiments.

UBS also promotes a wider perspective: seeing ethics as part of the behavioral sciences. Ethical theory tends to look inward to its own tradition, so that attempts to use empirical work narrow their focus to the assumptions of ethical theorists (Doris & Stich 2005).

2. Poor competitors. Ethics has long defined itself in opposition to social science. Gintis's proposed UBS makes a more difficult foe for those who define ethics in simple contrast of "ought" and "is," that is, prescriptive and descriptive. Some earlier candidates for core models in the social sciences, self-interested rational choice and selfish gene sociobiology, made this reactive account of ethics too easy. Gintis shows just how easy with a famous example:

> That human nature is selfish was touted as a central implication of rigorous biological modeling. In *The Selfish Gene*, for example, Richard Dawkins asserts that "[w]e are survival machines – robot vehicles blindly programmed to preserve the selfish molecules known as genes. . . . Let us try to teach generosity and altruism, because we are born selfish" (Dawkins 1976, p. 7). (Target article, sect. 10, para. 2)

Similarly, rational choice theory (RCT), in contrast to Gintis's proposed beliefs, preferences, and constraints (BPC) model of agency, is also an easy foil for ethics. RCT, as a powerful normative theory, should be a strong competitor to ethics. Unfortunately, the common assumption of self-regard makes it easy for ethicists to reject economics as too narrow. In both cases the field of ethics could purchase distinctiveness easily: ethics becomes anti-economics or anti-egoism.

Note that Gintis does not reject egoism in either biology or ethics for moral or theoretical reasons. He argues that the issue turns on empirical claims of behavioral game theory: "The result of the laboratory and field research on strong reciprocity is that humans indeed often behave in ways that have traditionally been affirmed in sociological theory and denied in biology and economics" (sect. 10, para. 6).

3. Constructing ethical science. UBS offers new methods that promise to improve the quality of work in ethics, by moving from the often theory-bound debates that characterize much of ethics to a tighter coupling of models and experiments.

A good example is the way one of Gintis's recent projects was driven by an experimental anomaly. An ultimatum game experiment in a small-scale society produced unexpectedly non-prosocial results. Henrich's "Machiguenga outlier" led to the widely cited cross-cultural experimental project in 15 cultures reported in Henrich (2004). To take another example, the role of reciprocity in ethics is understudied, yet both theory and experimental evidence indicate that human populations consist mainly of reciprocators of various kinds (Kurzban & Houser 2005). Third, Bicchieri (2006) provides a model of social norms that relates their descriptive and normative roles and accounts for much of the relevant experimental evidence.

This stress on experiments will surprise those who see ethics as too complex for experimental methods. On the contrary, a unified behavioral science sees experiments as crucial just for this reason:

> Without experiments, it is difficult to choose among the many possible hypotheses. In particular, anonymous one-shot experiments allow us to distinguish clearly between behaviors that are instrumental towards achieving other goals (reputations, long term reciprocity, and conformance with social rules for expediency sake) and behaviors that are valued for their own sake. (Henrich 2004, p. 10)

More generally, Gintis's UBS provides a broad foundation for a naturalized ethics. He provides evolutionary evidence for "prosocial traits, such as empathy, shame, pride, embarrassment, and reciprocity, without which social cooperation would be impossible"[1], as well as neuroscientific evidence for "both the neural plasticity of and the genetic basis for moral behavior" (sect. 5, para. 10).

ACKNOWLEDGMENTS

Thanks to Nick Wright for helpful discussion and Geoff Petrie for comments on a draft.

NOTE

1. We should not overplay the unity in "unified"; see the sharp debate between Binmore (2006) and Gintis (2006a) over the interpretation of some of these results.

Game theory for reformation of behavioral science based on a mistake

DOI: 10.1017/S0140525X07000672

Jeffrey Foss

Department of Philosophy, University of Victoria, Victoria, British Columbia, V8W 3P4, Canada.

jefffoss@uvic.ca

http://web.uvic.ca/philosophy/aboutus/faculty.php

Abstract: Gintis assumes the behavioral (=social) sciences are in disarray, and so proposes a theory for their unification. Examination of the unity of the physical sciences reveals he misunderstands the unity of science in general, and so fails to see that the social sciences are already unified with the physical sciences. Another explanation of the differences between them is outlined.

Gintis's ambitious theory faces tremendous odds. Expressed simply, it is revealed as a manifesto for the reformation of the behavioral (=social) sciences: evolution is the ultimate cause, gaming the proximate cause (sect. 1); accepting this will unify the social sciences. Like reformers before him, he castigates the status quo: the social sciences "offer partial, conflicting, and incompatible models. . . . it is scandalous that this situation was tolerated throughout most of the twentieth century" (sect. 12, para. 1).

So Gintis is game (in one colloquial sense of the word): plucky, spirited, showing fight. If his program for unification of the social sciences were to succeed, he would join the ranks of a tiny number of justly famous visionaries (only two examples leap to mind): Newton (who unified celestial and terrestrial physics) and T. H. Huxley (who unified botany, zoology, evolutionary theory, biochemistry, and microbiology in his textbook [See Huxley & Martin 1888/1875] to create biology). My assessment is that Gintis's vision has a number of blind spots, which when filled in do not look at all like he supposes. I will sketch in just one of them here: the present unity of the

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

physical sciences. If my sketch is accurate, Gintis's program is game in another (colloquial) sense: lame (as in having a game leg).

Gintis speaks blithely of the "seamless integration of physics, chemistry, and astronomy, on the basis of a common model of fundamental particles and the structure of space-time" (sect. 1, para. 2). Where is biology? It is a physical science that accepts fundamental particles and space-time, so why is it left out? We will return to this question below. First, let us consider the unity of the physical sciences in general. This unity has five dimensions (Foss 1995; 1998), as follows.

1. Ontology. Physics sets the ontology of science as a whole; it tells us what the primary constituents of reality are and their properties. Anyone who would plausibly call herself or himself a scientist must accept that the world and everything in it is composed of fundamental particles in space-time moved by the fundamental forces.

BUT, *all* scientists, whether physical or social, *already* accept this ontology. Gintis's above-quoted words show he conceives of scientific unity ontologically. So, his efforts simply are not needed. Psychologists, economists, and so on, accept the ontology he moots – but recognize this ontology is insufficient to explain the phenomena in their fields. Psychologists know that the physical mechanism, the brain, is made of fundamental particles, and economists know that the social mechanism, money, depends on these particles (in particular those composing our brains). But this knowledge hardly gets them to square one in the quest to explain the phenomena they study. So shared ontology is not the sufficient condition, the mother lode, of unification Gintis assumes it to be.

2. Explanation. Newton's equations explained not only projectile motion on Earth, but the motion of the Moon and planets, as well, effectively uniting two previously separate fields: terrestrial and celestial mechanics. Explanatory unification of (apparently) disparate phenomena still obtains today. For example, vision, radio, medical x-rays, and lightning are united by the electromagnetic equations which explain and predict their behavior.

BUT, it would be rash to think that this sharing of explanatory models and resources obtains across the physical sciences as a whole. Indeed, it does not obtain even within the paradigmatically unified science, physics. Famously, quantum phenomena are explained by different principles than are relativistic phenomena. Moreover, quantum theory and relativity theory have resisted all attempts at unification in a single theory: they are, in Gintis's terms, incompatible. Yet physics is unified. More familiarly, buoyancy, torque, levers, and a host of other purely physical phenomena are explained without any reference to the fundamental theories of physics Gintis cites. As we move from physics to chemistry, biochemistry, microbiology, and biology, the fundamental particles and forces become more and more irrelevant, and explanation must turn to the properties of the assemblies and mechanisms these particles and forces make possible. So the explanatory use of varying, incompatible, theories does not entail disunity, and the social sciences are not disunified on this basis.

3. Method. In one sense all scientists are united by the same method: close observation of natural phenomena, creation of explanatory theories and models, testing these theories and models, and so on. In another sense their disparate methods define specific scientific specialties: The physicist trains instruments and intelligence on heavenly bodies; the chemist trains different instruments and intelligence on terrestrial bodies, and so on.

BUT, social scientists offer no exception to this pattern, either among themselves or among scientists in general.

4. Shared information economy. A psychologist uses an electron microscope to see the microstructure of a neuron, and in so doing accepts in good faith what the physicist says about the microscope's reliability. The physicist likewise accepts the psychologist's model of the brain as a naturally occurring biological computer, and accepts in good faith that the mind is physical. Thus, their work and theories mutually reinforce each other by exchange within their shared information economy. More generally, information is the coin of the scientific realm, and its exchange unites the scientific community.

BUT, the social sciences are clearly in this economy, just like the physical sciences.

5. Sociology. The scientific community is organized to support its information economy, and to protect and improve the scope and accuracy of the information therein. Its institutions protect this economy by regulating membership in the community and deterring counterfeit information.

BUT, social scientists are bona fide members of this community, just like physical scientists.

So Gintis's diagnosis of disarray in the social sciences is misconceived, and his program based on a mistake. His disappointment with the social sciences is understandable, however. Our desire (indeed, need) to understand ourselves is extreme, and remains largely unsatisfied. Why? Because we are the most complicated phenomenon science has yet confronted. Protons and stars are simpler by orders of magnitude than a single living cell. This is why biology does not appear on Gintis's list of physical sciences: Biology deals with precisely the same complex phenomena that exceed the explanatory resources of physics in the social sciences – but the fact that biology spans the gap between them is further evidence of their continuity. The physical sciences, moreover, have a three-century head start on the social sciences. Better, then, to let the latter get on with its work, rather than force it to follow misconceived images of the former.

In evolutionary games, enlightened self-interests are still ultimately self-interests

DOI: 10.1017/S0140525X07000684

Thomas Getty

Department of Zoology and Kellogg Biological Station, Michigan State University, Hickory Corners, MI 49060.

getty@msu.edu

http://www.msu.edu/user/getty/index.htm

Abstract: Evolutionary theory provides a firm foundation for the unification of the behavioral sciences, and the beliefs, preferences, and constraints (BPC) model is a useful analytical tool for understanding human behavior. However, evolutionary theory suggests that if other-regarding preferences expressed by humans have evolved under selection, they are ultimately, if not purely, in the constrained, relative self-interests of individuals who express them.

Herbert Gintis is a distinguished economist with an unusually firm grasp of evolutionary theory and an unusually large number of entertaining book reviews posted at Amazon.com (87, as of this writing; I give the collection five stars). These reviews frame the issues in Gintis's target article quite nicely. In his review of *Vaulting Ambition: Sociobiology and the Quest for Human Nature*, by Philip Kitcher, Gintis (2005c) describes how "Edward O. Wilson's great work *Sociobiology* unleashed a furor of vitriolic criticism from mainstream social scientists." Wilson defined sociobiology as the systematic study of the biological basis of all social behavior, and he suggested that the social sciences were "the last branches of biology waiting to be included in the Modern Synthesis" (Wilson 1975, p. 4). The proclamation that the social sciences are branches of biology was not universally well received. Wilson's Figure 1.2 showed sociobiology engulfing surrounding fields like an amoeba engulfs its prey. No wonder there was such a fuss! In his review of *The Company of Strangers: A Natural History of Economic Life*, by

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Paul Seabright, Gintis (2005a) says: "Despite the rough treatment handed to Edward O. Wilson's call for a unification of biology and the social sciences some three decades ago, . . . the process of integrating social science into natural science appears to be in full swing." In his review of *Defenders of the Truth: The Battle for Science in the Sociobiology Debate and Beyond*, by Ullica Segerstråle, Gintis (2000b) concludes: "The sociobiologists and behavioral ecologists won the scientific war." If the war is over, Gintis's efforts here (in the target article), to unify the behavioral sciences under evolutionary theory in general, and the beliefs, preferences, and constraints BPC model in particular, can be seen as a postwar reconstruction effort: framework building. I like the framework Gintis proposes, but I doubt that the war is over (and I am betting that some of my fellow commentators will confirm this).

From my perspective as a behavioral ecologist, Gintis is mostly preaching to the choir. Evolutionary biology forms a foundation for all behavioral disciplines. Gene-culture coevolution, evolutionary game theory, and the BPC model should be important components of a general framework for understanding the social behavior of humans. Gintis offers the useful analogy: "just as physical principles inform model creation in the natural sciences, so must biological principles inform all the behavioral sciences" (sect. 2, para. 2).

What does it mean to have your research "informed" by fundamental laws, but not "reduced" to those laws? I wish Gintis had pursued this. He calls it scandalous that the behavioral disciplines have partial, conflicting, and incompatible models, but he does not get around to showing how useful the proposed unified theoretical framework could be. That was a missed opportunity. Interesting perspectives on how behavioral ecology has benefited from a firm foundation in evolutionary theory are provided by Krebs and Davies (1997b), McNamara et al. (2001), and Owens (2006). Once you understand evolutionary theory, and are comfortable with its limitations, you are bound to find it useful. Here is one way: It is a peculiar feature of science that we have such elaborate rules for how to test hypotheses, but so little guidance on how to create them. At the least, the deductive framework provided by evolutionary theory would allow social scientists to spend less time sorting through myriads of creative but short-lived, ad hoc hypotheses.

I like Gintis's proposed unifying framework, but I disagree with his interpretation of other-regarding preferences, including strong reciprocity. In his target article, Gintis quotes R. D. Alexander, from his book *The Biology of Moral Systems* (1987, p. 3): "ethics, morality, human conduct, and the human psyche are to be understood only if societies are seen as collections of individuals seeking their own self-interest." Gintis then introduces the concept of enlightened self-interest, without explaining his thoughts on the Alexander quote. They are clear from his review of Alexander's book at Amazon.com (Gintis 2000a), where the quote is followed immediately by this:

> This is of course the model of human action in standard economic theory, and I have spent my whole life dealing with its inadequacies. . . . Alexander's description of human behavior ignores strong reciprocity (spontaneously contributing to social goals and punishing shirkers and other non-contributors when there is no reasonable probability of future payoffs for the individual. (Gintis 2000a)

In his target article, Gintis sticks to the argument that morality nd costly other-regarding preferences have evolved, even though "it is implausible to expect that these costs will be repaid." This does not make sense in the evolutionary calculus. If costly other-regarding preferences have evolved in response to selection, then somehow or another they are ultimately in the constrained, relative self-interests of the individuals who express these traits, at least in the kinds of social environments where these other-regarding preferences would have been selected (Nakamaru & Iwasa 2006; West et al. 2006). There may be some confusion arising from inconsistent interpretations of the meaning of self-interest. Evolutionary biologists adopt an inclusive fitness perspective and include the welfare of kin or kin groups as being in the selfish interest of an actor (Axelrod 1981). Hagen and Hammerstein (2006) provide a critique of Gintis's interpretation of the seemingly selfless behavior of human subjects in contrived experimental games.

There are many important unanswered questions about the evolution, mechanisms, and dynamics of other-regarding preferences, which are clearly important in human social behavior. I agree with Gintis that "the notion that preferences are purely self-regarding must be abandoned" (sect. 12, para. 2) by anyone who might have this notion. However, the hypothesis that other-regarding preferences have been selected because they are ultimately in the self-interests of those who express them neither assumes, nor implies, that they are "purely" self-regarding. In his review of *The Evolution of Morality (Life and Mind: Philosophical Issues in Biology and Psychology)*, by Richard Joyce, Gintis (2006b) says, "A moral sense helps us be reasonable, prosocial, and prudential concerning our long-term interests." This seems like a sensible hypothesis to me. In evolutionary games, long-term and enlightened self-interests are still ultimately self-interests.

ACKNOWLEDGMENTS

I thank R. D. Alexander for getting me excited about this subject and Brian Abner for stimulating discussions of the issues. This is Kellogg Biological Station contribution number 1270.

Diversity, reciprocity, and degrees of unity in wholes, parts, and their scientific representations: System levels

DOI: 10.1017/S0140525X07000696

Robert B. Glassman

Department of Psychology, Lake Forest College, Lake Forest, IL 60045.
glassman@lakeforest.edu
http://campus/lakeforest.edu/~glassman/

Abstract: Though capturing powerful analytical principles, this excellent article misses ways in which psychology and neuroscience bear on reciprocity and decision-making. I suggest more explicit consideration of scale. We may go further beyond gene-culture dualism by articulating how varieties of living systems, while ultimately drawing from both genetic and cultural streams, evolve sufficiently as unitary targets of selection to mediate higher-level complex systems.

How best to understand wholes and parts? Nonliving systems' components hold their positions obediently, but components of living systems have endogenous spontaneity, so sometimes jostle each other for a larger share of their synergy. The stressed whole may then either decay or find a strong new shape in which to persist for another while.

Reciprocity and agency. Moving from physical sciences to biobehavioral sciences, *agency* appears, as the emergents acquire autonomy (Glassman 2006). Their cybernetic processes internally comprise regulatory feedbacks, while game theory becomes relevant to their external interactions. Agents play games. They do so in proliferating ways, from coalitions and competitions within cells to those of individuals, nations, and civilizations. "Strong reciprocity" (sect. 10) adds potential for a general theory of *levels*, concerning factors enabling groupings to become sufficiently unitary to emerge as elementary components of "higher-level" systems. However, theorizing is impeded in this excellent target article's excessive emphasis that adaptiveness focuses on individuals or genes. Additionally, the article misses aspects of psychology and neuroscience.

Metonymic fallacies in psychology. Labels can lead to exclusionary imperialism. We should disallow co-option of the

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

term "decision making" (sect. 3) by a subfield. All psychology concerns decisions. For example, textbooks of developmental psychology (e.g., Siegler et al. 2003) describe how children take another's point of view. An introductory psychology textbook (Weiten 2007) explains how perceptual cues feed decisions about objects' distances. Social psychology investigates seductive ways persuaders elicit disproportionate reciprocations (Cialdini 2001); relatedly, the applied psychology of marketing taps biases in purchasing decisions, using "segmentation and positioning" (e.g., Schiffman & Kanuk 2004). Evolutionary pressure to limit alternatives in decision-making may underlie the smallness of working memory capacity (Glassman 2000; 2005).

Although behaviorism continues to be iconic of academic psychology's self-conscious positivism, in Gintis's article it becomes a straw man. The universe of psychology has expanded, yet we should remain wary of its statistical homogenizations. For example, the section 6 discussion of "conformist transmission" would benefit from explicit coverage of diverse forms and their contingent relationships.

Brain idolatry: Localization can be a distraction. The target article's discussion of the "fitness-enhancing" decision-making ability of human "complex brains" (sect. 3) contains the homunculus fallacy. Such statements, as in the last paragraph of section 5 on prefrontal cortical substrates for moral behavior, merely add natural science "sanctity" to behavioral science. Neuroscience findings must be examined in conjunction with additional knowledge from behavioral sciences, humanities (including literature, even theology; Glassman 2004), and our human intuitions. Contrary to section 4, neuroscience does *not* imply a rational actor; that appearance arises from the researchers' presuppositions about what to study.

Why are people "faulty logicians" (sect. 9.6)? The discussion of the work of McClure et al. (sect. 4) stopped short of saying how neuroscience might elucidate mental processes (rather than merely ratify their existence), but it provides an opening: Suppose natural selection – parsimoniously feeling out costs and benefits – has really yielded only two brain systems for weighing short- versus long-term payoffs. Is there an "engineering limit" to how well those systems "overlap in the middle"? Are they disadvantaged with certain input parameters, though satisficing under most circumstances? If so, physiological parametric studies would cast light on our psychology.

Levels, or scales, of scientific investigation. Behavioral scientists should stop thinking strabismally about genetic and cultural information streams (Glassman et al. 1986). We do this even having amply acknowledged the "Promethean fire" (Lumsden & Wilson 1983) of coevolution. Behavioral science needs additional levels populated by representations of entities and processes having their own robustness. A fertile behavioral science will acknowledge a great variety of targets of selection, or "replicators" (sect. 2). For example, if discussion of "social capital" (Bowles & Gintis 2005b) were further developed in the "framework for unification of the behavioral sciences," it might better acknowledge that *expansions* of populations comprise different strategies than those underlying *longevity*.

Applied behavioral science: A critical example-problem challenge. Today, strife occurs with accelerated history in a globalized world shrunken for better and worse by technologies. The news of world-threatening violence in the Mideast so incredibly exists beside the historical fact that a thousand years ago Baghdad was an intellectual and spiritual center of Judaism, in harmony with high Muslim civilization. A millennium and a half earlier, the Iranian Cyrus helped Jewish contributions thrive (e.g., Johnson 1987; Konner 2003). What sorts of "games" have civilized peoples played on historical scales beyond our horizons of ordinary thoughtful vision? Though politically incorrect, we must ask: Who *are* these people? Who *were* we? In what ways have our genetic and cultural streams flowed down the ages, coalescing, diverging, coalescing and diverging again, from our living, loving, and hating bodies and our created artifacts then, to now? What motivations and emotions engaged at times and places of want or affluence? What were the mediating replicators and routes of reciprocity?

Equifinality, loose coupling, and persistence in cooperation and conflict. Living systems can "do the same thing in different ways," called *equifinality* by general systems theorists (Laszlo 1973; von Bertalanffy 1968). Among the examples is "motor equivalence" of intended actions (Lashley 1930; Milner 1970, p. 67). This suggests great variability in routes of reciprocation by which parts of a system support one another. Thus, living systems show *persistence*, even though in close-up, any pair of mutually supporting parts is *loosely coupled* (Glassman 1973). Evolution and history are the stories of shared functions and of oppositional tensions ranging from mutually regulatory to destructive.

Do systems ever "fail for success"? What prevents a stasis of the whole, regulatory parts harmonizing perfectly? (For example, do human beings really want the Garden of Eden without compelling, destabilizing temptations?) On the other hand, what keeps any system's dynamic subsystems-within-subsystems from crumbling to mere entropic Brownian wiggling of smallest elements?

Is there – without a teleological *deus ex machina* bolting down from the sky – a "meta-regulation" for viability? Do all subsystems somehow "seek" degrees of dynamism that keep one leg always over the boundary of predictability, vulnerably out in a realm of fundamental risk? How might such a thing work?

Perhaps inner levels progressively adjust to reflect the outermost. This hypothesis implies gradual "downward causation" (Campbell 1974), feedthrough from "upper" to "lower." For example, levels of the central nervous system from brainstem to cortex have accommodated one another while evolving (Rosenzweig et al. 2005; Striedter 2005), always with limited plasticity (Glassman & Smith 1988).

Evolutionary stages become intermediaries: "Progress." Notwithstanding social scientists' aversion to the term, perhaps there is "progress." In considering the peculiar suppleness of evolving systems, William F. Wimsatt (1980; Glassman & Wimsatt 1984) picked up Herbert Simon's allegory of two watchmakers – the wise Hora, who built interchangeable subassemblies in intermediate stages, and the shortsighted Tempus, who assembled everything in one fell swoop from 1,000 parts – if fortunate not to be interrupted. Although the criticism of teleology is allayed in remembering that an apparent immortally progressive evolutionary line is but one of many radiations, some longer-lived than others, intermediate stages do seem to evolve greater potency or resilience in becoming new levels of parts within wholes of complex systems.

The longer the world goes on, the more variety it witnesses. Even so, the identification of generalities such as strong reciprocity suggests that behavioral science need not be plagued interminably by relativistic perspectivism (Krantz 2001); unity may be achieved.

ACKNOWLEDGMENT

Conversations with Hugh Buckingham concerning an article we are coauthoring about the history of neuroscience have improved my understanding of levels and of unity in science.

Do the cognitive and behavioral sciences need each other?

DOI: 10.1017/S0140525X07000702

David W. Gow, Jr.

Cognitive/Behavioral Neurology Group, Massachusetts General Hospital, Boston, MA 02114.

gow@helix.mgh.harvard.edu

Abstract: Game theory provides a descriptive or a normative account of an important class of decisions. Given the cognitive sciences' emphasis on

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

explanation, unification with the behavioral sciences under a descriptive model would constitute a step backwards in their development. I argue for the interdependence of the cognitive and behavioral sciences and suggest that meaningful integration is already occurring through problem-based interdisciplinary research programs.

Gintis's proposal is not for everyone in the behavioral sciences. This is clearly the case for cognitive psychologists. Language processing, memory, problem solving, categorization, and attention are not easily construed as instances of strategic interaction. Gintis is aware of this and notes that some of us will have little to gain from the current proposal. I would argue that, in fact, psychologists and other behavioral scientists have a lot to lose by adopting it.

Gintis's inspiration, the unification of the physical sciences in the twentieth century, provides important perspective on the problem. Prior to Niels Bohr and Ernest Rutherford's elucidation of atomic structure in 1912, chemistry was a highly developed but largely descriptive discipline. Sophisticated bench techniques had been developed, and Mendeleev's periodic table of 1869 stood as a towering intellectual accomplishment. Nevertheless, chemistry was primarily a descriptive field concerned with the cataloging of elements, reactions, and material properties. Atomic and then quantum theory transformed chemistry from a descriptive to an explanatory science, opening new vistas including molecular biology, atomic physics, and quantum chemistry.

Game theory, though a powerful tool with potential broad application in the behavioral sciences, offers cognitive psychologists a different transformation when considered as a unifying theory for the behavioral sciences. Even overlooking the limits of its application and granting it some of the elegance of Mendeleev's periodic table or the computational precision of Keppler's descriptions of planetary movement, we are left with the fact that game theory is ultimately a descriptive or a normative tool. While it may describe an important class of human decisions, it does not provide sufficient insight into the mechanisms that produce our decisions. Gintis expresses surprise at the fact that cognitive psychology devotes most of its energies to understanding "the processes that render decision making possible" (sect. 3, para. 3); but of course, this is exactly what we must do if we are to truly understand those decisions. The evolution from explanatory to descriptive science would be a great step backwards. The natural consequence of such a choice would be the formation of a deepening fissure between the cognitive and behavioral sciences. I would hate to see that happen, and suspect that it never will.

A fissure between the cognitive and behavioral sciences would be undesirable for everyone involved. Consider the fundamental but challenging problem of framing in the beliefs, preferences, and constraints (BPC) model. Gintis notes that subjects show framing biases because they tend to map the formal structures of games encountered in the lab to experiences or facets of their normal lives. As a cognitive psychologist I would argue that the framing bias reflects limits imposed by operating characteristics of human memory, attention, and problem-solving, as well as the way that listeners map linguistic descriptions of task parameters onto conceptual representations. Rather than dismissing all deviations from the predictions of the model as "performance errors" (sect. 3, para. 4), game theorists could improve their models by addressing how cognitive mechanisms produce systematic variation in performance. The cognitive sciences also need to explore the domains occupied by the behavioral sciences to explore the full complexity of how different representational types and processes interact over time in the broader social context that defines human experience. This recognition of a broader context also enables the realization of the cognitive science's potential for practical application.

If the BPC model, or at least the BPC model as it is currently envisioned, is not the right kind of theory to truly unify the behavioral sciences, what are behavioral scientists to do right now? With a widely recognized need for integration, researchers in a variety of fields have staked out an integration strategy that I suspect will preserve the traditional bonds between cognitive and behavioral sciences. Rather than weaving an encompassing theory to unify the sciences, many researchers have adopted a piecemeal strategy of using insights and techniques from allied fields to constrain research and theory development.

Experience in spoken language communication demonstrates the advantages and disadvantages of this approach. Historically, the sequence of effects that constitutes language communication is often conceived as a chain involving discrete processes including motor planning, sound production and filtering, auditory transduction, phonetic classification, and lexical access. Each of these steps has been adopted by a different set of specialists with expertise in areas ranging from physics to physiology and psychology. Over the last 50 years, this strategy has often been productive, in part because it has broken a challenging problem into more manageable ones, and it has allowed researchers to draw on well-developed theory and techniques from the mother disciplines of each specialty. These advances have come at a cost, though. Considered together, this approach forms a kind of production line in which the output representations of each step (e.g., acoustic structure, auditory representation, phonetic representation, etc.) form the input representation for the next. In Marr's terms (1982), these input and output representations are a critical defining characteristic of the computational problem. But problems arise when practitioners in one field adopt a shallow or outdated view.

For example, early speech perception research suggested that listeners show strict categorical perception of speech sounds. This view was adopted by linguists in part because it resonated with their existing theories, and by psychologists who used this insight to frame spoken word recognition as the process of mapping between discrete abstract representations of phonetic categories and similarly abstract representations of word forms. Unfortunately, psychologists and linguists failed to absorb the full complexity of the original data or appreciate the implications of subsequent data showing that discrete categoricality is largely an artifact of metalinguistic task demands and that categorization is a gradient process that may be strongly influenced by context. As Gow (2003) notes, this failure led word recognition theorists to overlook the dynamic properties of lexical activation and their role in specific processing challenges including lexical segmentation and lawful phonological variation. Happily, interdisciplinary efforts by phoneticians, psychologists, and linguists have begun to address the full complexity of phonetic categorization processes in the context of word recognition. The lesson here is that cross-disciplinary integration is not impossible in principle. It simply requires training and scholarship that are defined by problems rather than historical disciplinary boundaries. Ideally, this would strengthen work within disciplinary boundaries by giving researchers additional perspective on the assumptions, techniques, and theories that define them.

As the recent explosion of neurocognate fields (e.g., neuroeconomics, social neuroscience, neuroethics) attests, researchers in many of the disciplines are turning to interdisciplinary neuroscience approaches to ground behavioral science theories. Gintis takes this tack himself, citing neuroscience evidence to ground claims about the representation of payoff, hyperbolic discounting, and the basis for moral behavior and cultural replication. Cognitive neuroscience, as it emerges from its current, largely descriptive phase, provides an obvious potential source of grounding for the behavioral sciences. Whether or not it ultimately provides the grand theory that unifies the behavioral sciences will likely depend on the degree to which researchers in the neurocognate fields are able to develop solid, current problem-based expertise in both neuroscience and individual disciplines within the behavioral sciences with an eye towards explanation.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Gintis meets Brunswik – but fails to recognize him

DOI: 10.1017/S0140525X07000714

Kenneth R. Hammond

Psychology Department, University of Colorado, Boulder, CO 80309.
kenneth.hammond@colorado.edu **Brunswik.org**

Abstract: With a few incisive (and legitimate) criticisms of crucial experiments in psychology that purported to bring down the foundations of modern economics, together with a broad scholarly review that is praiseworthy, Gintis attempts to build a unifying framework for the behavioral sciences. His efforts fail, however, because he fails to break with the conventional methodology, which, regrettably, is the unifying basis of the behavioral sciences. As a result, his efforts will merely recapitulate the story of the past: interesting, provocative results that are soon overthrown because they are limited to the conditions of the experiment. Gintis is keenly aware of this limitation – and thus meets Brunswik, but fails to recognize him; this we know because he seems unaware of the fact that Brunswik said it all – and provided a detailed defense – a half century ago.

All previous attempts to build a unifying framework in psychology have failed because the scientific base keeps collapsing. What makes this lamentable *is that Gintis's effort is so unnecessary*; yet Gintis has done it again. That is, he has removed scientific achievements of the past by pointing out methodological mishaps, without recognizing the fundamental ever-present nature of the methodological flaw that causes the mishap. (See Hammond and Stewart [2001] for a history of previous attempts.) Gintis's attempt may even be unique; he forms a framework that rests on foundations that he himself tears down. Although he builds his framework on achievements in various fields with remarkable expertise and admirable scholarship, he must qualify them all by noting that they are restricted to the artificial conditions under which they were obtained; as he puts it, they may not hold in "real life" or in the "real world." (These phrases are important to him; he uses them at least nine times.)

A key example is Gintis's dismissal of Tversky, Slovic, and Kahneman's foundational demonstration of preference reversal in the choice of lotteries (Tversky et al. 1990), by simply pointing out that "the phenomenon has been documented only when the lottery pairs A and B are so close in expected value that one needs a calculator... to determine which would be preferred by an expected value maximizer" (sect. 9.2, para. 5). And "when the choices are so close to indifference, it is not surprising that inappropriate cues are relied upon to determine choice" (sect. 9.2, para. 5). So there it is; in those few sentences, Gintis dismisses 30 years of the celebration of a major finding by psychologists that trumped (or so they believed) a major underpinning of economic theory and led to a Nobel Prize.

But 50 years ago, Brunswik pointed out psychology's vulnerability to this kind of attack when he said: "Generalization of results concerning ... the variables involved must remain limited unless the range, but better also the distribution ... of each variable has been made representative of a carefully defined set of conditions" (Brunswick 1956, p. 53). In short, specific values of the independent variable should be chosen by demonstrably defensible relations to a "carefully defined set of conditions" representative of the universe to which the results are intended to apply. That is what Brunswik meant by "representative design"; and that is what Gintis wants, but unfortunately doesn't know it.

But he is not alone; like Gintis, neither do Tversky, Slovic, and Kahneman want to bother with this admonition; for them it is sufficient to arbitrarily select (or were these the only values that produced the desired effect?) certain data points and ignore the entire question of generalization. Tversky et al. were willing to be arbitrary in their choice of data points; Gintis will be satisfied with a wave to the "real world," no specification necessary. But it is precisely that arbitrariness that brings down massive generalizations. In short, had Tversky, Slovic, and Kahneman listened to Brunswik (1956), they would not have exposed themselves to a fatal criticism that Gintis, at least, believes destroys a cornerstone of their work. Of course, they were not alone; had psychologists in general been willing to listen to Brunswik (which they were not), we would have been spared the spectacle of Gintis's scholarship taking down that piece of the foundation that psychologists had found so important.

Yet Gintis goes on to the destruction of his own foundation. He makes use of Brunswik's principal admonition (without mentioning him) of the impossibility of generalizing results from "rule of one variable" experiments to what Gintis calls "real life." That gap between experimental conditions and conditions outside the experiment rules out the possibility of the generalization Gintis needs in order to establish his claim of meaningfulness for research in game theory and the like.

Hence, Gintis fails to explain just how to get rid of the "artificiality" that he claims undermines the psychological experiments because he doesn't know how. Nor would he know how to get rid of the arbitrariness in Tversky et al.'s choice of data points in the independent variable that defeats their purpose. He could, of course, suggest other data points; but without paying heed to Brunswik's admonition to make them representative "of a carefully defined set of conditions," they would be just as arbitrary as those chosen by Tversky et al.

This is strange; all that needed to be done was to cite Brunswik's words that I have cited above (that is the easy part) and then act accordingly; that is – *justify, theoretically or empirically*, your choice of data points on the independent variable! That is the hard part. Few mainstream psychologists will acknowledge the point from Brunswik (1956) that I cited earlier, which bears repeating, that "generalization of results concerning ... the variables involved must remain limited unless the range, but better also the distribution ... of each variable, has been made representative of a carefully defined set of conditions" (p. 53). Few will go to the trouble of specifying "a set of conditions" toward which their results are intended to apply, and then accept that "the range, but better also the distribution... of each [independent] variable" should be justified. Psychologists have been turning their heads from this troublesome matter for 50 years. As a result, an economist can dispense with a body of research with a few simple sentences.

In short, Gintis's bold effort to provide a "unifying framework" will be frustrated because a "unifying framework" *already exists* in the behavioral sciences. That unifying framework is the behavioral sciences' ironclad commitment to a methodology that prevents valid generalization, and which, therefore, leads some of its most scholarly members to bemoan the gap between artificial experiments and – forgive us – the "real world." Brunswik never put it better than when he said: "There is little technical basis for telling whether a given experiment is an ecological normal, located in a crowd of natural instances, or whether it is more like a bearded lady at the fringes of reality, or perhaps a mere homunculus of the laboratory out in the blank" (1956, p. 204). So, was the Tversky et al. (1990) experiment, so often quoted and relied upon in countless textbooks and professors' lectures, "more like a bearded lady at the fringes of reality," or did it reflect the circumstances we are interested in? Who can tell?

Brunswik could, and did in 1956. Sadly, few have listened. Gintis shows us the consequences. And may learn from them himself.

Rationality versus program-based behavior

DOI: 10.1017/S0140525X07000726

Geoffrey M. Hodgson

Department of Accounting, Finance and Economics, The Business School, University of Hertfordshire, Hatfield, Hertfordshire AL10 9AB, United Kingdom.
g.m.hodgson@herts.ac.uk **http://www.geoffrey-hodgson.info**

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Abstract: For Herbert Gintis, the "rational actor," or "beliefs, preferences, and constraints (BPC)," model is central to his unifying framework for the behavioral sciences. It is not argued here that this model is refuted by evidence. Instead, this model relies ubiquitously on auxiliary assumptions, and is evacuated of much meaning when applied to both human and nonhuman organisms. An alternative perspective of "program-based behavior" is more consistent with evolutionary principles.

Herbert Gintis makes a powerful case for a unifying framework for the behavioral sciences and I agree with much of it. One of my principal concerns is his elevation of the "rational actor," or "beliefs, preferences, and constraints (BPC)," model. At the outset he admits that the term "rational" has "often misleading connotations" and states a preference for the "BPC" description. However, the word "rational" or its derivatives appear in the article more often than the "BPC" term. Gintis spends much of the article refuting claims in the literature – particularly from psychology and experimental economics – that the assumption of rationality is refuted by the evidence. Other important features of his unificatory framework, such as the "evolutionary perspective" and "gene-culture coevolution" are given much less overall attention.

Gintis abandons many of the established meanings of rationality, including the narrower version in which individuals are motivated by their self-interest. For him, the rational actor model "depends only on choice consistency and the assumption that an individual can trade off among outcomes" (sect. 4, para. 2). I do not argue that choice consistency (otherwise known as preference transitivity) is refuted by the evidence. Instead, I uphold it would be difficult in practice to find any evidence strictly to refute this assumption.

An experiment may seem to reveal preference intransitivity, by showing that while X is preferred to Y, and Y is preferred to Z, Z is preferred to X. However, this can be explained away by showing that the three pairwise comparisons did not take place under identical conditions, or were separated in time or space. The consumer could have "learned" more about her true tastes and expectations during the experiment itself, or other factors may account for the apparent intransitivity. All we have to do is show that the two Zs in the above comparisons are not quite identical. They could be slightly different in timing, substance, or their informational or other contexts. We then get the result: X is preferred to Y, Y is preferred to Z_1, and Z_2 is preferred to X. In these circumstances, transitivity is no longer violated. In short, preference inconsistency is extremely difficult to detect in practice because it is impossible to replicate the strictly identical global conditions under which choice rankings are made.

When a proposition is difficult or impossible to falsify, then we should worry, even if we do not uphold the strict Popperian criterion of falsifiability as the mark of science. As recognized in the philosophy of science (Nagel 1961), a problem with non-falsifiable propositions is that they are consistent with any conceivable evidence in the real world, and hence their explanatory power is diminished (Hodgson 2001; Udéhn 1992).

However, this does not mean that the rational actor framework is necessarily useless or wrong. Gintis can point to many examples of its apparent success, not only in economics, but also in biology, sociology, and elsewhere. In the face of this apparent success I have a different claim: In every case, the results of such models depend critically on assumptions that are additional to typical axioms of rationality. For example, Gary Becker contended that standard rationality assumptions generate a number of predictions concerning human behavior. However, all of Becker's claimed "predictions" depend on assumptions *additional* to his core axioms of rationality with given preferences (see, e.g., Becker 1981). Indeed, because it is difficult to conceive of evidence that falsifies these axioms, such models must depend on auxiliary assumptions in order to generate specific results (Blaug 1992, p. 232; Vanberg 2004).

Although Gintis treats it as equivalent to rationality, the "beliefs, preferences, and constraints" phraseology evokes a different set of concerns. The problem, particularly from an evolutionary perspective, is explaining where beliefs and preferences come from. How do they evolve? From an evolutionary perspective, they can no longer be taken as given.

Just as the meaning of "rationality" is undermined when it is applied to all organisms simply on the basis of the existence of consistent behavior, Gintis evacuates the term "belief" of much of its meaning when he suggests that it applies to all organisms. Do bacteria have "beliefs" in the same sense that humans have "beliefs"? No. Bacteria lack deliberative, linguistic, communicative, prefigurative, and other capacities that humans use to construct beliefs. All organisms that possess a developed nervous system have some kind of neural activity, but that does not mean that it involves "beliefs." The danger is that the terms "beliefs" and "reasons" become so broad that they include habits, emotions, instincts, and visceral reactions. The BPC/rationality model becomes so capacious that it accommodates any form of impulse towards behavior, deliberative or otherwise.

Consequently, we should not identify "rationality" as a supreme overarching principle for the behavioral sciences. There is an alternative, with an even stronger evolutionary grounding. What are common to all organisms are not beliefs but *behavioral dispositions*. These come in two kinds. First, there are genetically inherited dispositions of instincts. Second, there are dispositions that are acquired during development and interaction with the environment; these are known as habits, and are most important in social animals with imitative capacities and complex brains. Nevertheless, the capacity to acquire habits itself requires instinctive priming. Furthermore, rationality, in the more meaningful sense of conscious rational deliberation, also depends on habits and instincts as props (Hodgson 2004; Plotkin 1994).

Both instincts and habits are rule-like dispositions: In circumstances X, the organism strives to do Y. Sets of rule-like dispositions are linked together into what we may term *programs*. The biologist Ernst Mayr (1988) argued for an alternative general behavioral perspective along these lines. Instead of simply assuming that agents hold beliefs and preferences, the paradigm of program-based behavior ties in with an explanation of their evolutionary emergence, through both natural selection and individual development. Evolution involves both the adaptation of programs to changing circumstances and the elimination of other programs through selection. Whereas the rational actor model simply sets out assumptions that are consistent with behaviors, the paradigm of program-based behavior focuses on the explanation of the dispositions behind any act. The concept of the program may be subdivided between programs that do and do not involve deliberation or conscious prefiguration, thus avoiding the dangerous conflation of these different meanings with the misleading terminology of rationality.

The paradigm of program-based behavior has been applied to economics by Viktor Vanberg (2002; 2004) and has strong similarities with John Holland's (1995) theory of adaptive agents. Both in terms of its terminology and its focus of explanation, it is more general than the rhetoric of rationality and beliefs. A danger with this rhetoric is that it extends concepts with a special and largely human meaning to a broader context, and thereby denudes them of much substance.

Implications for law of a unified behavioral science

DOI: 10.1017/S0140525X07000738

Owen D. Jones

Departments of Law and Biological Sciences, Vanderbilt University, Nashville, TN 37203-1181.

owen.jones@vanderbilt.edu
http://law.vanderbilt. edu/faculty/jones.html

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Abstract: The argument for unifying behavioral sciences can be enhanced by highlighting concrete, vivid, and useful benefits that coherent behavioral models could provide. Shifting sets of behavioral assumptions underlie every legal system's efforts to modify behaviors by changing incentives in the legal environment. Consequently, where those assumptions are flawed, improved behavioral models could improve law's effectiveness and efficiency in regulating behavior.

It has become common, fortunately, for scholars to call for greater interdisciplinarity, believing important syntheses will follow. Suggestions on *how* to synthesize are much rarer. It is therefore significant that Gintis does not only argue eloquently that the "abiding disarray in the behavioral sciences" (sect. 7, para. 5) should be overcome through unification. He also proposes how we might achieve unification: by using evolutionary theory and game theory to bridge disciplines. I suggest that the appeal of this argument can be argumented by more explicit and extensive attention to the practical advantages of unifying, and of unifying the particular way Gintis proposes.

We know inconsistency is costly. Some people (and universities) are wasting their resources developing theories that cannot be right. Unification (which itself is costly) could yield net gains by reducing that waste and offering new and larger benefits in its place. These benefits obviously include coherence, as it is elegant and satisfying to integrate perspectives into a seamless whole. And coherence helps generate new and more accurate knowledge, a worthy end in itself. But beyond this, highlighting practical *applications* of a unified model will help underscore the need for coherence.

Concrete benefits of the kind of unification Gintis calls for can be illustrated by considering the legal system. At its most basic, the legal system is a massive, intricate, and dynamic set of tools for changing, channeling, and regulating the behavior of humans. Put simply, society uses law to manipulate environmental conditions in ways that prompt more of the individual and group behaviors we want, and less of the behaviors we don't. For example, society uses law to: facilitate economically efficient exchange (by enforcing bargains); protect private property from theft; stem aggression; force expansions or contractions of healthcare coverage; protect people from unsafe or ineffective drugs; allocate rights and duties; prompt suitable savings rates for retirement; and regulate sexual, mating, and reproductive behavior – to name but a few.

But with the rare exception of when law physically forces a behavior (e.g., through arrest and incarceration), anything significant that law achieves it achieves by *incentivizing* people to behave differently (through taxes, fines, rewards, threats, and the like). This means that at the core of every legal policy is an implicit behavioral model that provides the fulcrum for the lever of law (Jones 2004). That fulcrum contains the shifting set of assumptions that underlie a prediction: if law moves this way, behavior will move that way, as intended, and not some other way.

This reality sharply clarifies the practical need for the unified behavioral model Gintis helps prompt us to construct. As a consumer and end-user of behavioral models, and charged by society with accomplishing very specific behavioral changes, law has a particularly acute need for improved models of human behavior. For law can obtain no more leverage on human behaviors than the solidity of this fulcrum affords. That is, inaccurate assumptions make for soft fulcra and poor leverage. Or, put another way: Incorrect assumptions about human behavior impede law's ability to effectively and efficiently change behaviors, and to do so at the least cost to society. Unfortunately, law's existing models are outdated. Legal thinkers generally overrely on social science perspectives and insufficiently attend to conflicts among them, or between them and life science perspectives. The generally unarticulated assumption is that all law-relevant behavior arises exclusively through environmental and cultural pathways.

The flaw in this assumption, and the weakness it creates within the fulcrum for legal action, can be remedied in part by the kind of integration of evolutionary thinking for which Gintis calls. To be concrete, evolutionary analysis in law can help us among other things to: discover useful patterns in behaviors that law regulates; uncover conflicts between two or more existing policies; sharpen cost-benefit analyses; deepen our understanding of how the human animal behaves in law-relevant ways; provide theoretical foundation and potential predictive power; disentangle multiple causes and the policies that conflate them; expose unwarranted assumptions underlying legal policies; better assess the comparative effectiveness of legal strategies; and reveal deep patterns in legal architecture (Jones & Goldsmith 2005).

So, for example, where Gintis predicts that unification will increase our understanding of law-abiding behavior and the dynamics of criminality (among many other things), I believe he is demonstrably right, and that such understanding can potentially translate into concrete gains in different areas of applied behavioral science, such as law. Some existing work (see the bibliography at the *Society for Evolutionary Analysis in Law* website: www.sealsite.org) helps illustrate his point. Incorporating evolutionary analysis into legal thinking may help law to more effectively combat child abuse and rape (Jones 1997; 1999). It may help law devise optimal incentives (O'Hara 2004). It predicts that main features of legal systems (such as property; Stake 2004) will tend to reflect the effects of natural and sexual selection on the main features of the evolved human neural architecture (Jones 2004, p. 1706). It predicts that both human morality (Alexander 1987) and human intuitions about just punishments will reflect evolved sensitivities concerning bodily harm, resources, and exchanges (Hoffman & Goldsmith 2004; Robinson et al., in preparation). It predicts that many of the human irrationalities that pose practical problems for law (such as endowment effects, overdiscounting of future interests, and spiteful behavior) may flow from evolutionary causes and reflect a mismatch (a "time-shifted rationality"; Jones 2001) between ancestral and modern conditions (Gigerenzer 2002). And it predicts a "law of law's leverage," whereby the effectiveness of a given unit of legal intervention will vary with the extent to which an evolved predisposition toward now unwanted behaviors was adaptive for its bearers, on average, in past environments (Jones 2001).

This is just a beginning. Were the kind of unification for which Gintis calls to be more aggressively pursued, as I believe it should be, it could yield practical gains analogous to those that Darwinian medicine may yield (Nesse & Williams 1994/1996). Specifically, a successful unification of the human behavioral sciences could improve the legal system's effectiveness and efficiency in regulating human behaviors. Making this case could help make the argument for unification even more appealing than it already is.

Disciplinary stereotypes and reinventing the wheel on culture

DOI: 10.1017/S0140525X0700074X

David P. Kennedy
RAND Health, RAND Corporation, Santa Monica, CA 90407.
davidk@rand.org
http://www.rand.org

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Abstract: Gintis argues that disciplinary models of human behavior are incompatible. However, his depiction of the discipline of anthropology relies on a broad generalization that is not supported by current practice. Gintis also ignores the work of cognitive anthropologists, who have developed theories and methods that are highly compatible with the perspective advocated by Gintis.

Gintis's argument that disciplinary models of human behavior are incompatible relies on broad generalizations of disciplinary models that are not empirically justified. Gintis's treatment of anthropology is especially thin and describes a stereotypical view of *culture* that is not used by contemporary anthropologists to a significant degree.

Despite the centrality of the concept of culture to anthropology, there is currently no consensus about what culture is within anthropology, not even within the subdiscipline of cultural anthropology. One branch of anthropology, cognitive anthropology, has already offered a view of culture similar to the one advocated by Gintis (D'Andrade 1995). For the remainder of this review, I will summarize a definition of culture used by cognitive anthropologists in order to argue that much of the work done towards unifying disciplinary views of human behavior and culture has already been done by cognitive anthropologists.

Despite disagreements about the nature of culture, there is agreement that culture refers to something learned rather than inherited (Brumann 1999). The concept of culture is usually invoked to understand the behavior and thought patterns of groups. However, only individuals can learn, and they are the only source of cultural data (Handwerker 2001). Therefore, any definition of culture must begin with the knowledge that human beings possess, and how individual human beings learn and process information. Because culture is learned primarily through other people, it is also the result of social interaction and is shared. This results in culture being both socially and individually constructed.

The way individuals construct cultures begins with the formation of cognitive models of reality. Humans are limited in their ability to recall discrete units of information (Miller 1956). However, humans have an almost unlimited ability to "chunk" together bits of information into schematized models of particular domains of information (D'Andrade 1995). There is not a one-to-one correspondence between a particular model and a particular domain. Rather, multiple models are at work in concert at any given time. Some models are more likely than others to be invoked at a given moment because of a weighting process that develops over time after repeated experiences with a domain (Strauss & Quinn 1997). This weighting process is mediated by emotions that are evoked during these experiences. Models invoked during experiences that are associated with positive emotional feedback are more likely to be used in the future. The opposite can be said for negative emotional feedback (Strauss & Quinn 1997). Cognitive anthropologists define the complete set of an individual's cognitive models, including the models' associated emotional weights and behaviors, as the raw material of culture (Handwerker 2001).

Instead of a unitary, internally consistent "seamless web" that contains unambiguous rules for behavior (DiMaggio 1997), cognitive anthropologists see culture as fragmented and inconsistent. At any point in time, individuals may have internalized cognitive models that are contradictory. These models, although guides for behavior, can never have a one-to-one correspondence with behavior outputs because of their heterogeneity. Rather than acting as a blueprint for behavior, culture acts like a "toolkit" of strategies which individuals use to choose among behavioral options depending on momentary external circumstances (DiMaggio 1997).

The cognitive models that individuals have at their disposal at any point in time develop as a result of past experiences and are constantly being modified with new experiences. Cognitive models influence the behavioral choices that individuals are forced to make in the context of external circumstance. These behavioral choices then provide individuals with additional experiential information from which ideas and emotions are subsequently generated and modified. Because no two individuals have exactly the same experiences, no two individuals have the same set of cognitive models. And no one person has the same set of cognitive models from one moment to the next, because individuals are constantly behaving and processing additional experiential information (Handwerker 2001).

Although individuals are the only source of cultural data, and the raw materials of culture pertain to individuals, culture is created through social interaction. "Cultural models" refers to models that are to some extent shared by members of a population (Dressler & Bindon 2000). However, since cognitive models are the result of a creative process within individual brains, culture is not a "thing" that can be transferred from one person to another (Handwerker 1989). Because an individual's set of cognitive models is the end product of life experience, and because members of populations often have similar, if not identical, experiences, this produces patterning of cognitive, emotional and behavioral traits. Also, as individuals interact with members of their social networks, they experience the world vicariously through those other network members. This enables individuals to hold ideas and emotions about experiences and behaviors without actually experiencing them directly. Therefore, the cognitive models and emotions of each individual human being depend in some part on the cognitive models, emotions, and behaviors of other members of their social networks.

Individual human beings do not passively accept models from their social network. Rather, they accept the models that work, modify those that do not, and "share" these modifications back into their social network in a dynamic, continually evolving creative process. When models developed through previous experience are unable to account adequately for new stimuli, individuals switch from "automatic" to "deliberative" cognition, which they use to actively and innovatively restructure their own models to better account for new stimuli (DiMaggio 1997). Subsequent interaction with a social network leads to the spread of the innovation throughout the network if the innovation is successful at resolving similar inadequacies in the models held by other network members (Tomasello 1999). The "spread" of innovations throughout a network is actually individual brains making similar cognitive adjustments after interactions with members of their social networks. Thus, culture can be shared, but only metaphorically and imperfectly (Handwerker 1989).

This view of culture is highly compatible with Gintis's objectives. This decades-old tradition of scholarship based on the findings of cognitive science and centered around the collection of cultural data should be considered before the reinvention of the wheel.

The flight from reasoning in psychology

DOI: 10.1017/S0140525X07000751

Joachim I. Krueger

Department of Psychology, Brown University, Providence, RI 02912.

joachim@brown. edu

http://www.brown.edu/Departments/Psychology/faculty/krueger.html

Abstract: Psychological science can benefit from a theoretical unification with other social sciences. Social psychology in particular has gone

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

through cycles of repression, denying itself the opportunity to see the calculating element in human interaction. A closer alignment with theories of evolution and theories of interpersonal (and intergroup) games would bring strategic reasoning back into the focus of research.

Gintis observes the incompatibility of the multitude of mini-paradigms in the social and behavioral sciences and judges this state of affairs to be scandalous. His argument is that if there are so many incompatible paradigms, many of them must be wrong. This may be so, but it does not represent the worst possible state of affairs as long as some of these paradigms are correct. If the many paradigms were replaced by a single one, and that one turned out to be false, the damage would be great indeed. To move toward a unification of the social and behavioral sciences, Gintis proposes a "take-the-best" heuristic that recombines paradigm fragments that have proven empirically useful and that are compatible with one another. This will not amount to a true scientific revolution *sensu* Kuhn (1962), because that would require an entirely new look at the whole field and an overthrow of the dearest theoretical assumptions across the board.

As a psychologist, I agree with Gintis's claim that psychology must shed its distrust of reasoning, and especially strategic reasoning in social contexts. To avoid the topic of thinking is no way to resolve the rationality question. Every generation of psychologists seems to reclaim the irrelevance of reasoning using the tools of the day. First came the idea that if rats and pigeons can be trained to perform complex behaviors, parsimony demands that complex human behaviors be explained by animal learning models (Skinner 1971). Then came the idea that social behavior is "unbearably automatic" (Bargh & Chartrand 1999). Unbearable indeed. The idea that higher reasoning can be dismissed because some critical behavior can be elicited in the laboratory without the participant's awareness is the logical fallacy of affirming the consequent. Finally, the current rush toward neuroscience is yet another flight from reasoning (Kihlstrom 2006). Despite its undeniable scientific interest and importance, brain imagery can reveal only correlates of reasoning, not reasoning itself.

Why does reasoning have such a bad name in psychology? One consideration is that strategic reasoning implies the ability to outthink and deceive others. The capacity of research participants to be one step ahead mentally is always a concern in the laboratory. To allay this concern, experimenters seek ways to circumvent strategic reasoning, and then mistake what is left for the whole of psychology. A related consideration is a common misunderstanding of the relationship between determinism and human choice. The point of strategic reasoning is to be unpredictable when so desired. Yet, when determinism is taken to entail predictability, unpredictable behavior seems undetermined, and therefore either random or "freely willed." The implication of free will and the reference to intentions or desires seems like a throwback to Aristotelian thinking, according to which the apple falls to the ground because it wants to.

I believe these worries are ill-founded. Even in a fully deterministic world, strategic reasoning can occur. Perhaps such reasoning is unpredictable in principle, much like the nonlinear mathematics of chaos theory, or it is just sufficiently unpredictable by those conspecifics it is designed to deceive. If it is the latter, its purpose is served, and we can get on with the task of modeling it. Likewise, intentions need not be mere by-products created by brains that are *really* only in the business of generating behavior (Krueger 2003). Recent advances in neuroscience show that tetraplegics can be fitted with prosthetic devices that receive neural signals associated with conscious intentions and translate them into motor behavior (Hochberg et al. 2006).

In his effort to build a comprehensive "model of individual human behavior," Gintis has surprisingly little to say about how strategic reasoning can retake center stage. As he notes, however, the study of "rationalizability" is one place to begin. True, with enough assumptions, almost any behavior may come to appear reasonable. What is needed is a compass that helps chart a course between unprincipled post hoc rationalization and the equally barren strategy of demonstrating irrationality with experimental designs that equate any significant finding with the presence of a bias or an error (Krueger 1998).

Many social-psychological phenomena that presumably illustrate the fallibility of social behavior and cognition can be rationalized with the tools of decision analysis or game theory. To illustrate, consider the classic finding of bystander apathy (Darley & Latané 1968). The more potential helpers there are, the less likely is an individual to assist a person in need. Orthodox social-psychological analysis focuses on victims facing life-and-death emergencies and bystanders who have little to lose by helping. However, a full model requires the bystanders' costs and benefits, as well as the number of bystanders, to be variables.

Game theorists have derived precise predictions for behavior in the volunteer dilemma. A person caught in this dilemma hopes that others will bear the cost of intervening, but would intervene herself if she knew that no one else will. According to one solution (Diekmann 1985), a bystander will help with a probability of

$$1 - \left(1 - \left[\frac{1}{N} \times \frac{c}{b}\right]^{(1/(N-1))}\right).$$

Notice that this probability becomes smaller as the cost of helping, c, or the group size of bystanders, N, increases, and as the benefit to the helper, b, decreases. This is a mixed-motive solution that maximizes the expected value for the bystander. Incidentally, this solution also predicts Darley and Latané's (1968) finding that a victim becomes slightly more likely to receive aid from *someone* as the group becomes larger.

Other classic and contemporary findings can be rationalized along similar lines. Gintis's emphasis on the evolutionary rationality of conformity and imitation is another good example. Although the reorientation of the social and behavioral sciences proposed by Gintis may not (yet) amount to a Kuhnian revolution, it may turn out to be a decisive first step to overcome disciplinarian parochialism. We can begin today by reading – at least from time to time – one another's journals.

The limitations of unification

DOI: 10.1017/S0140525X07000763

Arthur B. Markman

Department of Psychology, University of Texas, Austin, TX 78712.

markman@psy. utexas.edu

http://www.psy.utexas.edu/psy/faculty/markman/

Abstract: There are two roadblocks to using game theory as a unified theory of the behavioral sciences. First, there may not be a single explanatory framework suitable for explaining psychological processing. Second, even if there is such a framework, game theory is too limited, because it focuses selectively on decision making to the exclusion of other crucial cognitive processes.

Can the behavioral sciences be unified? The target article suggests that it is critical to develop a single theoretical framework that can be used to explain phenomena across the behavioral sciences and to develop new questions. The article

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

correctly notes that when any science defines its theoretical constructs narrowly with respect to particular phenomena, it may miss key generalizations across situations. Chemistry would be limited indeed if it had separate theories for each element. Within psychology, there has often been a tendency to define psychological processes with respect to particular experimental tasks (e.g., attention, categorization, decision making), and to develop separate theories for each (Uttal 2001). Therefore, the call to look across phenomena to develop key principles for understanding behavior is a welcome one.

Despite my general enthusiasm for taking a broad view of behavioral phenomena, I address two concerns in this commentary. First, are there likely to be a small number of unifying theories for the behavioral sciences? Second, to what degree is game theory a good candidate for a unifying theory?

How many theories? Theories are powerful things. They determine the questions that scientists find interesting to answer and the types of data that scientists collect. Therefore, a theory chosen to unify a set of disciplines must match the phenomena being studied if it is to be useful. The target article argues, for example, that the lack of a unifying theory in the behavioral sciences lowers the credibility of the sciences. In the Middle Ages, however, the alchemists had a unifying theory that merged physical and spiritual issues. In retrospect, that theory was not credible, because the physical and spiritual do not have a common cause.

It is an empirical question whether there is sufficient causal similarity in the explanatory factors across disciplines in the behavioral sciences to warrant a small number of unifying theoretical frameworks. Unlike the objects of study of physics and chemistry, behavior is a result of a long series of adaptations. As the target article points out, some of the functions of mind are rooted in evolutionary adaptations to particular environments, but others result from a long coevolution of the human organism and human culture.

This combination of evolution and culture has likely programmed humans with many different and distinct mechanisms that enable complex behaviors. At this stage, it is not clear that a single framework can be developed that will encompass all of these functions. To demonstrate, I consider a few of the limitations of game theory.

Game theory as a unifying framework. The target article assumes that decision making is the core of psychological processing. This assumption is critical for the use of game theory as a unifying theory. At one level, it is possible to view human behavior as a set of choices, because any behavior by an individual or group is one of many behaviors that could have been carried out, and the selection of a particular behavior can be modeled as a choice.

However, this view of behavior will serve to illuminate key questions only for behaviors that relate explicitly to choice. In cases in which people are not explicitly choosing, game theory might be consistent with behavior, but it will not explain how that behavior comes to be. To the extent that a science cares about individual behavior, it will have to have theories that unify only those psychological processes that have a common basis.

As one example, the cultural transmission of information takes place via communication processes that do not seem obviously explicable within game theory. The target article recognizes this gap and draws on the notion of a "meme" in an attempt to relate communication to evolution (Dawkins 1976). The concept of a meme is an interesting metaphor for communication drawn from evolutionary theory, but it is hardly a viable theory of cultural transmission of ideas (Atran 2001). At best, it describes some factors that lead ideas to be preserved from one individual to another. A theory of linguistic and nonlinguistic communication will be an integral part of theories in disciplines including psychology, communications, education, and anthropology.

As a second example, theories of behavioral science have to take the concept of mental representation seriously (Markman 1999). What people are able to represent about their environment determines what can be communicated via culture (Medin & Atran 2004). Furthermore, the representations people form affect their ability to satisfy their basic needs, including obtaining food and recognizing other members of their species (Hirschfeld 1996). Game theory assumes that people have mental representations, but it does not provide significant constraints on the nature of those representations. Because questions of mental representation do not fall naturally out of formulations of game theory, research driven by this framework is likely to gloss over issues of representation.

I close this commentary with an analogy: Human cognition may be more like computer programming than it is like physics or chemistry. A successful computer program is one that works. Although there are guidelines for how to write programs, and endless debates over what programming languages are best, there is no unified theory of programming (Hayes 2006). What will work best depends on the computing environment, the problem to be solve, the available hardware, and the other programs with which a particular piece of software must integrate. Similarly, human cognition is a collection of routines. Some of these are based on the biological niche occupied by humans and human forebears. Some are the result of culture. Some are the result of learning by individuals as a result of life experience. Theories of human behavior must recognize the diversity of functions that the cognitive system supports. As the target article assumes, decision making is crucial to survival. It is not the sole cognitive function, however, and cannot serve single-handedly as a unifying framework across all of the behavioral sciences.

Probabilistic equilibria for evolutionarily stable strategies

DOI: 10.1017/S0140525X07000775

Roger A. McCain

Department of Economics and International Business, Drexel University, Philadelphia, PA 19104.

mccainra@drexel.edu **http://william-king.www.drexel.edu/**

Abstract: This commentary suggests that an equilibrium framework may be retained, in an evolutionary model such as Gintis's and with more satisfactory results, if rationality is relaxed in a slightly different way than he proposes: that is, if decisions are assumed to be related to rewards probabilistically, rather than with certainty. This relaxed concept of rationality gives rise to probabilistic equilibria.

Gintis's target article concedes that the rational-action core of game theory will be a difficulty for many scholars. On the whole, Gintis's strategy is to introduce beliefs as an autonomous factor in decisions along with preferences and constraints, and to suggest that well-known empirical anomalies in rational action theory can be isolated as errors in beliefs. Gintis goes further in relaxing the rational-action model, suggesting that Nash equilibrium is too narrow and that the broader game-theoretic concept of *rationalizability* is sufficient for his purposes. However, there is a difficulty here that suggests a logical inconsistency, in that an evolutionarily stable strategy, a central concept in evolutionary game theory, is a Nash equilibrium that satisfies some other conditions, as well. Rationalizability is applicable to one-off play in which there is no repetition or learning, whereas evolutionary game theory is largely based on models of repeated matching and can be a model of social learning.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

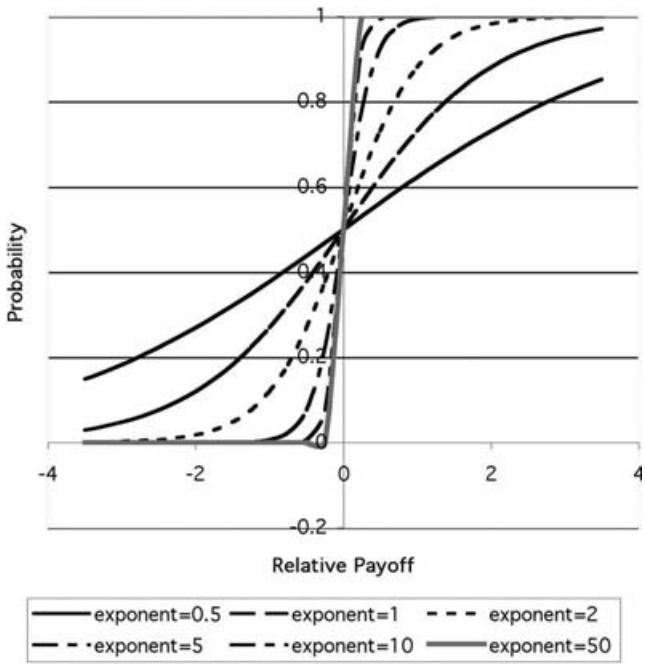


Figure 1 (McCain). Probability of the choice of strategy A.

This comment suggests that the equilibrium framework may be retained, with far more realistic results, if rationality is relaxed in a slightly different way: that is, if decisions are assumed to be related to rewards probabilistically, rather than with certainty. This relaxed concept of rationality gives rise to probabilistic equilibria (e.g., Chen et al. 1997; McKelvey & Palfrey 1995).

Suppose that an agent is to choose between two courses of action, A and B, where B pays zero and the payoff to A varies from -4 to $+4$. Suppose then that the probability that the agent will choose strategy $i = \mathrm{A}$, B is given by $\mathrm{P}_i = \mathrm{Y}_i^{\theta}/\sum_j \mathrm{Y}_j^{\theta}$ where Y_j is the payoff to strategy j. Then the probability that i is chosen increases with the relative payoff Y_i. This is shown in Figure 1 for several values of the exponent. As Figure 1 suggests, the exponent theta can be thought of as an index of relative rationality, in that the choice of the higher-payoff strategy is more probable, on the whole, when theta is larger.

If we consider a game-like interaction between two or more agents, each must consider the strategy choice of the other as a probability distribution and base his own choice of strategies on the expected values of payoffs from his own strategies. A probabilistic equilibrium then is a set of probability distributions over strategy choices that are mutually consistent in that each is an approximately best response to the other. Consider, in particular, the small centipede game shown in Figure 2. This game can also be represented in normal form, using the contingent strategies shown in Table 1. We can compute a probabilistic equilibrium for this game by numerical methods (McCain 2003). Assuming a value for theta of 2 (Anderson et al. 2004, p. 1044) and computing a probabilistic equilibrium based on the strategies in Table 1, we obtain the probabilities for the nine possible strategy

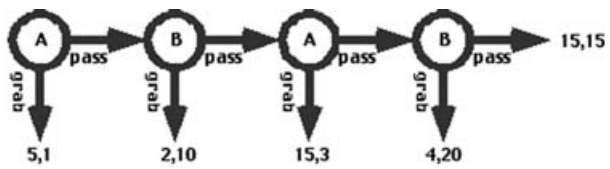


Figure 2 (McCain). A small centipede.

Table 1 (McCain). *Contingent strategies for the small centipede*

	Al
1	grab
2	pass, and if Bob passes once, grab
3	pass, and if Bob passes once, pass again
	Bob
1	If Al passes once, then grab
2	If Al passes once, then pass and, if Al passes twice, then grab
3	If Al passes once, then pass and, if Al passes twice, then pass

Table 2 (McCain). *Probabilities of the strategy combinations in an example of probabilistic equilibrium in the centipede*

		Al		
		1	2	3
Bob	1	0.655	0.152	0.018
	2	0.077	0.018	0.002
	3	0.062	0.014	0.002

Table 3 (McCain). *Probabilities of the strategy combinations in an example of probabilistic equilibrium in the centipede with reciprocity*

		Al		
		1	2	3
Bob	1	0.036	0.404	0.237
	2	0.009	0.105	0.061
	3	0.008	0.088	0.052

combinations as shown in Table 2. Note that the probability of a simple noncooperative equilibrium is about 65% in this example.

Gintis stresses the importance of non-self-regarding motives. Although there is little precedent in the literature, it is quite simple in principle to introduce non-self-regarding motives into a probabilistic equilibrium model. We need not be concerned whether a non-self-regarding act generates a "warm feeling" that increases the person's utility or not, nor whether people are in some ultimate sense self-interested even when their actions are non-self-regarding. We simply posit that the probability of choosing a strategy is influenced by some non-self-regarding considerations as well as the payoffs.

To continue with the example of the centipede game, suppose motives of reciprocity influence the probabilities of strategies in this case. To represent reciprocity we need some reference values, so that (for example) when an agent's payoff is less than the reference value he retaliates (negative reciprocity), and conversely. As Gintis stresses, these reference values may depend on social norms, but it is possible by

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

examining some games to make a plausible guess. In this case, assume that the reference payoffs are the payoffs the agent would get if he were to grab at his first opportunity; that is, 5 for Al and 10 for Bob. Whatever the probability of strategies 2,1 and 3,1 (Al passes and Bob grabs), this would give Al reason for negative reciprocity amounting to a shortfall of 3, and so reduce the probability of his choosing strategies 2 or 3. This is just one illustration. In general, we can indicate reciprocity by $(Y–Y_r)(Z–Z_r)$, where Y and Z are the payoffs to Al and Bob, respectively, and Y_r, Z_r are their reference payoffs. In place of Y_i^θ in the formula for the probability of strategy i, we write $Y_i^\theta + w[sigman(Y - Y_v)(Z - Z_v)] \sqrt{(ABS((Y - Y_v)(Z - Z)))}$[EQ03], where w is a non-negative weight representing the importance of reciprocity motives to the individual. Taking the exponent 2 as above and $w = 0.333$ for a single example, we have in Table 3 the probabilities for the nine possible strategy combinations. In this case, for example, we see the overall probability of Al choosing the cooperative strategy 3 is 0.35, by comparison with 0.022 in the previous case. Note that reciprocity can lead in some cases to multiple equilibria that reinforce both cooperative and noncooperative outcomes.

We see that the probabilistic equilibrium concept admits of a larger and more plausible range of outcomes in this case than Nash equilibrium does, particularly when non-self-regarding motives are introduced in a natural way. In summary, this conception of equilibrium has three major advantages in the context of Gintis's program:

1. It allows rationality to be a relative concept.
2. Although probabilistic equilibria for some games closely approximate deterministic Nash equilibria, in some other cases, including the centipede, they can be quite different and more plausible.
3. Non-self-regarding motives are easily introduced.

Extending the behavioral sciences framework: Clarification of methods, predictions, and concepts

DOI: 10.1017/S0140525X07000787

Alex Mesoudi[a] and Kevin N. Laland[b]

[a]*W. Maurice Young Centre for Applied Ethics, University of British Columbia, Vancouver, BC, V6T 1Z2, Canada;* [b]*School of Biology, University of St. Andrews, St. Andrews, Fife KY16 9TS, Scotland.*

mesoudi@interchange. ubc.ca
http://gels.ethics.ubc.ca/Members/amesoudi **knl1@st-and.ac.uk**
http://www.st-andrews.ac.uk/∼seal

Abstract: We applaud Gintis's attempt to provide an evolutionary-based framework for the behavioral sciences, and note a number of similarities with our own recent cultural evolutionary structure for the social sciences. Gintis's proposal would be further strengthened by a greater emphasis on additional methods to evolutionary game theory, clearer empirical predictions, and a broader consideration of cultural transmission.

Gintis presents a framework for the behavioral sciences that is both rooted in evolutionary theory and based around the "rational actor," or "beliefs, preferences, and constraints (BPC)," model of human decision making. We fully support many of Gintis's aims and themes – specifically, his attempt to integrate the diverse behavioral sciences around a common framework, basing that framework in evolutionary theory (both biological and cultural) within a gene-culture coevolutionary perspective, and advocacy of the use of simple mathematical models to explain complex real-world behavioral phenomena. Indeed, there are many parallels here with our own recent articles advocating integrating the social sciences around evolutionary theory (Mesoudi et al. 2004; 2006). We maintain that stronger parallels exist between biological and cultural evolution than generally recognized, and argue that the structural and theoretical bases of the social and behavioral sciences should be mutually consistent. As noted in Mesoudi et al. (2006), individual decision-making processes will have significant effects on cultural evolution; and as Gintis notes, *culture* – the body of knowledge, beliefs, attitudes, and norms that is acquired via social learning from conspecifics – will likewise have strong effects on individual decision making and behavior. However, we believe that there are opportunities to extend the specific framework proposed by Gintis, extensions that we regard as critical to the success of his venture. As it currently stands, we fear that only a minority of behavioral scientists would empathize with his stated objectives and methods.

First, although we welcome Gintis's emphasis on evolutionary game theory, and agree that there is considerable potential for further use of this powerful method throughout the human sciences, we fear that exclusively focusing on this method may be counterproductive. We see no reason for him to limit himself to just a single theoretical technique when others – such as population genetic models (Boyd & Richerson 1985; Cavalli-Sforza & Feldman 1981; Laland et al. 1995), agent-based simulations (Axelrod 1997b; Epstein & Axtell 1996; Kohler & Gumerman 2000), stochastic models (Bentley et al. 2004; Neiman 1995; Shennan & Wilkinson 2001), and phylogenetic methods (Lipo et al. 2006; Mace & Holden 2005; O'Brien & Lyman 2003) – may be more suitable in other cases. Not all problems in the human sciences can be treated as games; many are not even frequency dependent. For example, reconstructing linguistic phylogenies using cladistic methods (Holden 2002) is a valuable evolution-inspired analysis of a problem not amenable to evolutionary game theory. Although game theory might be ideally suited to certain problems in economics or political science, it is more difficult to envisage its use in, say, neuroscience. (The link Gintis makes between strategic interactions and mirror neurons [sect. 8, para. 6] is speculative at best.)

Second, Gintis's use of evolutionary reasoning at times appears somewhat empty and rhetorical, and we encourage him to elaborate on *how* he envisages researchers could use relevant biological knowledge. His proposals would benefit from the identification of the evolutionary processes responsible for the behavioral phenomena under discussion and the generation of clear predictions amenable to empirical testing, where such objectives are feasible. For example, Gintis argues that "the economist's model of rational choice behavior must be qualified by a biological appreciation that preference consistency is the result of strong evolutionary forces" (sect. 12, para. 2), without specifying exactly what these forces are. Presumably he means some form of selection; but the selection pressures need to be atypically well-established to allow us to predict the exact preferences, beliefs, values, and behaviors that people will acquire and in what contexts. If the exercise is not sufficiently constrained by scientific knowledge, there is a danger that it could degenerate into the idle speculation and weak inferences characteristic of some modern evolutionary psychology (Laland & Brown 2002). Similarly, the statement "the ease with which diverse values can be internalized depends on human nature" (sect. 12, para. 2) fails to specify exactly what is meant by "human nature," given that human behavior varies extensively over different times and in different contexts (Ehrlich 2000), and even unlearned predispositions may exhibit considerable variation within and between populations.

Third, we encourage further clarification and qualification of Gintis's use of the phrase "cultural transmission." Gintis seemingly overemphasizes conformist horizontal cultural transmission ("cultural transmission generally takes the form of

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

conformism" [sect. 1.1.2]) and mislabels other transmission biases as "conformist" (describing a strategy of "imitat[ing] what appear to be the successful practices of others" [sect. 6, para. 4] as an example of "conformist transmission," which it clearly is not). We agree that humans (and other animals) are often extremely susceptible to conformity effects, as shown by classic social psychology experiments (Jacobs & Campbell 1961; Sherif 1936). Nonetheless, the extent of conformist transmission in actual societies has yet to be empirically determined in sufficient detail, particularly relative to the numerous other social learning strategies that can be employed at different times and in different contexts (Laland 2004). Ethnographic studies often conclude that cultural transmission is vertical rather than horizontal (Hewlett et al. 2002; Ohmagari & Berkes 1997) or that considerable individual idiosyncrasy exists in culturally acquired beliefs (Aunger 2004), belying any strong conformity effect.

Many of the proposed problems with the BPC model and game theory discussed in section 9 may at least partially disappear if cultural transmission is more explicitly considered. Isolated individuals may be boundedly rational and exhibit various decision-making errors and biases, which often leads them to suboptimal behavior; but in large enough groups, a simple strategy of copying successful neighbors/behaviors can allow individuals to reach global optima quickly and cheaply. Recent experiments (Mesoudi & O'Brien, submitted) demonstrate that participants readily use and benefit from a copy-successful-individuals strategy, particularly in multimodal fitness landscapes. Gintis notes (sect. 9.6, para. 3) that scientists do not exhibit biases such as the representativeness heuristic; presumably this is because scientists have acquired the right solution (or the right means of solving such problems) from others, rather than having been born without such biases.

These criticisms aside, we endorse Gintis's scheme and hope that it provokes more interaction among psychologists, anthropologists, economists, and sociologists, as well as greater use of evolutionary theory within the human sciences.

Selection of human prosocial behavior through partner choice by powerful individuals and institutions

DOI: 10.1017/S0140525X07000799

Ronald Noë

Ethologie des Primates, DEPE-IPHC, CNRS and Université Louis-Pasteur, UMR 7178, Strasbourg, 67087 CEDEX, France.

ronald.noe@c-strasbourg.fr

Abstract: Cultural group selection seems the only compelling explanation for the evolution of the uniquely human form of cooperation by large teams of unrelated individuals. Inspired by descriptions of sanctioning in mutualistic interactions between members of different species, I propose partner choice by powerful individuals or institutions as an alternative explanation for the evolution of behavior typical for "team players."

I applaud Gintis for a brave and erudite attempt to unify the diverse and often contrasting approaches of the numerous scientific disciplines that meddle with human behavior. As a case in point he presents an array of explanations for seemingly "irrational" or "fitness reducing" contributions towards public goods shared with unrelated individuals. Altruistic punishment, an essential element of "strong reciprocity," seems Gintis's favorite example of a strategy with an irrational flavor (sect. 10. para. 5).

Here and elsewhere (Bowles & Gintis 2004a; 2004b; Gintis 2000d), Gintis endorses explanations for the evolution of such behavior in humans at two levels: (1) "classical" individual selection in repeated interactions among dyads, and (2) gene-culture coevolution with an essential element of selection at the group level. I am sympathetic to this line of thought, but nevertheless propose more reflection on other forms of cooperation before deciding that phenomena or evolutionary mechanisms require unique explanations. Cooperation can be found in a breathtaking number of forms in a wide range of organisms (Bshary & Bronstein 2004; Sachs 2004). I concentrate on cooperation between members of different species ("mutualisms," in ecological jargon), because here we also find both unrelated partners and multiple players.

Gintis points at the stabilizing effect of "punishment" in cooperating human groups (sect. 10, para 6). This reminds of a mechanism called "sanctioning" (Denison 2000; Kiers et al. 2003). Typically large and long-living individuals of one species dispose of mechanisms that allow them to reject the least profitable partners among many small partners belonging to a species with a short generation time. Examples are interactions between yuccas and yucca-moths (James et al. 1994; Pellmyr & Huth 1994) and between various plants and mycorrhizal fungi (Kummel & Salant 2006) or rhizobia (Simms et al. 2006). Sanctioning is an extreme case of "partner choice," which is a potent force of selection for cooperative behavior (Noë 2001).

There is an essential difference between punishment and sanctioning, however: selection for "selfish punishment" is favored when it changes the behavior of the punished to the benefit of the punisher. The same is true for "altruistic punishment," except that the benefit is shared by all members of the punisher's group, whereas the punisher ends up with a net cost. Punishing only pays because the punished individual is likely to interact with the punisher and/or his group again. A plant sanctioning unprofitable partners will typically not interact with the rejected individuals again. By cutting its losses, the plant obtains immediate benefits, which makes sanctioning akin to getting rid of parasites. Strong selection on the sanctioned partners is a by-product of this self-regarding strategy. The punishment concept stresses effects on individual behavior over intervals ranging from seconds to lifetimes. The sanctioning literature emphasizes the selective effect on sanctioned species at evolutionary time scales. Punishment can have a selective effect, as well, albeit probably weak, through its effect on the fitness of punished individuals. Rejection of partners in favor of more profitable ones, in turn, can shape the partners' behavior in favor of the chooser just like punishment, as long as the partners are not eliminated in the process.

Can plants sanctioning insects or fungi suggest explanations for the evolution of forms of cooperation that seem uniquely human? I concentrate on the selective force of partner choice, although selective pressure due to punishment should not be completely ignored either. A strong selective force exists when single powerful individuals (e.g., chiefs, kings, warlords, priests) or institutions (e.g., councils of elders) favor group members on the basis of their prosocial behavior. Such selection may take place during the formation of hunting parties, raiding teams, warfare, and the like. Partner choice can only be evolutionarily stable when it leads, as a rule, to an increase of fitness of both chooser and chosen. For example, the initiator of the hunt should be able to increase his individual returns by choosing the right hunters, and the participants should have higher benefits than those excluded from the hunt. This mechanism can contribute to the selection of altruistic behavior relevant to the production of public goods when partners are chosen on the basis of characteristics that make them good team players rather than good hunters per se: loyalty to the team, willingness to back up failing teammates, fairness in sharing, and so on. A recent study found that such traits are still highly relevant in modern societies: "Being a *team player* is of paramount importance in the workplace,

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

according to both employers and employees. Being perceived as a team player is considered to be more important than doing a good job, being intelligent, being creative, making money for the organization, and many other *good* qualities" (cited from the HOWFAIR study 2003 p. 1; Level Playing Field Institute 2006).

This partner choice scenario suggests stronger selection through partner choice in societies with stronger power asymmetries. A casual analysis of data presented in Henrich et al. (2004, Table 2.1 and Fig. 2.2) confirms this: The five ethnic groups classified as living in villages or chiefdoms act more altruistically than the five living in more egalitarian family-based societies (Mann-Whitney-U test $p < 0.05$).

However, teams without internal power differentials can act as powerful selective forces by recruiting new members even in egalitarian societies (Smith 2003). The uniquely human ability to report performances of team members to the rest of the community increases the necessity to acquire a good reputation as a team player and reinforces individual partner preferences. Some elements of cooperative behavior essential for the creation of public goods can also be selected in the context of dyadic interactions; for example, in trading and coalition formation. Finally, a selection for prosocial behavior is possible in the mating arena, as well: partners may be chosen, by their mates, their families, or their clans, on the basis of cooperative attitudes. Whatever form partner choice takes, I think it can rival with group selection as an explanation for typically human forms of cooperation and is better at explaining its selection at the genetic level.

Considering cooperation: Empiricism as a foundation for unifying the behavioral sciences

DOI: 10.1017/S0140525X07000805

John W. Pepper

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721-0088.

jpepper1@email.arizona.edu

http://eebweb.arizona.edu/Faculty/Bios/pepper.html

Abstract: Economics and evolutionary biology share a long history of interaction and parallel development. This pattern persists with regard to how the two fields address the issues of selfishness and cooperation. The recent renewed emphasis on empiricism in both fields provides a solid foundation on which to build a truly scientific unification of the behavioral sciences.

I applaud the goal of unifying the behavioral sciences. In considering what form such a unification should take, I will expand on an issue raised by the target article: the roles of empirical evidence and of long-established "models" or assumptions. In particular, the proposed unification should accommodate and extend recent parallel developments in how biologists and social scientists approach the issue of selfishness and cooperation, which is a central issue in both fields.

Apparently independently, both the social sciences and evolutionary biology have historically developed a strong assumption, verging on an ideological commitment, to the idea that all behaviors are fundamentally self-regarding. This assumption has been labeled by Henrich et al. (2005) as the "selfishness axiom." In recent years, however, there have been moves in both fields toward reconsidering this stance, based not on a priori assumptions or ideological commitments but instead on empirical evidence. These changes are very welcome, because they reaffirm the empiricism that lies at the core of all true science. Intentionally pursuing and expanding this return to empiricism, especially in approaching controversial topics, will provide a sound foundation on which to build the proposed unification of the behavioral sciences.

There is a long history of interaction, cross-fertilization, and parallel development between the fields of economics and evolutionary biology. If, as Gintis suggests, evolution is to provide a basic organizing principle for the behavioral sciences, recent parallel developments in economics and evolutionary biology are especially salient. What Henrich et al. (2005) call the "selfishness axiom" has long been a canonical model of human behavior for the social sciences. As Gintis notes, the idea that human nature is essentially selfish has also been touted as a central implication of Darwinian evolution. Because the theory of natural selection is based on competition for survival and reproduction, it has long been assumed that evolution through natural selection could only lead to selfish traits and behaviors. Similarly to the situation in behavioral sciences and especially economics, over the years this assumption ossified into something approaching an ideological commitment. This stands in marked contrast to Darwin's own approach, which always stressed the primacy of empiricism. Darwin (1859) anticipated Karl Popper's (1959) emphasis on falsifiability when he boldly stated, "If it could be proved that any part of the structure of any one species had been formed for the exclusive good of another species, it would *annihilate* my theory, for such could not have been produced through natural selection." (Darwin 1859/1966, p. 201, emphasis added).

Darwin's focus here on species rather than individuals is rather outmoded now, and the current version of the biological selfishness axiom is closer to that expressed by Williams: "[I]t should be possible to show that every adaptation is calculated to maximize the reproductive success of the individual, relative to other individuals, regardless of what effect this maximization has on the population" (Williams 1966, p. 160).

In biology, the selfishness axiom is based mostly on an appeal to intuition rather than on a rigorous and complete analysis, and more recent work has shown it to be somewhat oversimplified (Keller 1999; Sober & Wilson 1998; Wilson 2004). Nonetheless, when it was challenged in an even less rigorous way by Wynne-Edwards (1962), evolutionary biologists responded not merely with appeals to empiricism or to theory, but also with what amounted to a cultural taboo on the ideas of group-level selection and adaptation that Wynne-Edwards had espoused (Sober & Wilson 1998). The economics version of the "selfishness axiom" seems to have as little, or less, theoretical basis as the biological version, but rather to have arisen mostly as a cultural norm among economists.

The most unfortunate parallel between the two fields is that in each, poorly founded assumptions came to constrain empirical work by dictating that certain questions were uninteresting because the answers were already "known," or rather, assumed. It is encouraging that in both fields renegade workers have begun to ask unconventional questions and to receive unexpected answers, allowing the hypothesis-testing process of science to resume. As Gintis notes, recent experimental work by social scientists has shown that humans often behave in ways that have been traditionally denied in biology and economics. Indeed, the canonical model of *Homo economicus* now appears to have no known instantiation in any real human culture (Henrich et al. 2005).

In biology, key empirical breakthroughs have revealed that what we now recognize as (selfish) individual organisms originated as intensely and elaborately cooperative collectives. Proposed examples of such "transitions in individuality" include the origin of genomes and cells from replicating macromolecules (Eigen & Schuster 1978; Rohlfing & Oparin 1972); the origin of nucleated eukaryotic cells from bacterial associations (Margulis 1970); and the origin of multicellular organisms from single cells (Buss 1987). Now that we are aware of the enormous potential for ongoing conflict within organisms, both among cells (Buss 1987) and among genes (Burt & Trivers 2006; Hurst et al. 1996), the very existence of organisms that are well-integrated enough

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

to act, selfishly or otherwise, is a testament to the importance of cooperation in both the processes and the outcomes of evolution (Buss 1987; Maynard Smith & Szathmáry 1995/1997; Michod 1999; Ridley 2001).

If any field, including the study of behavior, is to lay claim to being a true science, it must without hesitation relinquish any assumptions or models that do not conform to ongoing and unbiased observations of empirical reality.

The integrative framework for the behavioural sciences has already been discovered, and it is the adaptationist approach

DOI: 10.1017/S0140525X07000817

Michael E. Price,[a] William M. Brown,[a] and Oliver S. Curry[b]

[a]*Centre for Cognition and Neuroimaging, School of Social Sciences and Law, Brunel University, Uxbridge, Middlesex UB8 3PH, United Kingdom;* [b]*Centre for Philosophy of Natural and Social Science, London School of Economics, London WC2A 2AE, United Kingdom.*

michael.price@brunel.ac.uk
http://people.brunel.ac.uk/~hsstmep/
william.brown@brunel.ac.uk
http://people.brunel.ac.uk/~hsstwmb/
o.s.curry@lse.ac.uk
http://www.lse.ac.uk/darwin

Abstract: The adaptationist framework is necessary and sufficient for unifying the social and natural sciences. Gintis's "beliefs, preferences, and constraints" (BPC) model compares unfavorably to this framework because it lacks criteria for determining special design, incorrectly assumes that standard evolutionary theory predicts individual rationality maximisation, does not adequately recognize the impact of psychological mechanisms on culture, and is mute on the behavioural implications of intragenomic conflict.

The unification of the behavioural sciences, and their integration with the rest of natural science, is currently taking place within a neo-Darwinian framework which views all organisms as bundles of adaptations (Tooby & Cosmides 1992). Gintis's "beliefs, preferences, and constraints" model (BPC) provides no convincing arguments for why it is a meaningful addition to the existing framework. Below we summarize why the existing framework is a necessary and sufficient one for unifying the disciplines.

An adaptation is a phenotypic device that was designed by selection to allow its encoding genes to outreplicate genes for rival devices. As the fundamental organisational principle of organismal tissue, adaptation is as indispensable to understanding human behaviour as it is to understanding any organismal trait. This does not mean, of course, that all traits are adaptations, but rather that, in order to understand organismal design, one must determine whether particular traits are adaptations, by-products of adaptations, or random noise. In order to establish that a trait is an adaptation, there must be evidence of special design (Williams 1966); that is, evidence that the trait was designed by selection for the specific purpose of solving a particular (set of) problem(s). Because BPC does not include criteria for testing for the existence of special design, it is often unable to determine whether a trait is an adaptation or not.

For example, Gintis notes that subjects in economic games often violate the predictions of traditional economic theory, and he concludes that their behaviour evidences an evolutionary process that favoured those who consistently behaved in individually fitness-damaging ways. However, people engage in many kinds of apparently fitness-damaging behaviours in novel environments (including experimental economic laboratories), and the observation of such behaviour is not a sufficient basis on which to conclude that the behaviour evolved for the purpose of producing a fitness-damaging outcome. It is as if, upon observing that many men spend significant time and money consuming pornography, thereby irrationally foregoing real mating opportunities, one were to conclude that a "preference" for pornography is the product of selection for fitness-damaging behaviour. However, pornography's popularity is more likely a result of semi-autonomous psychological mechanisms that evolved in a pornography-free world. Because there is no evidence that these mechanisms were specially designed for pornography, pornography's popularity is not evidence of selection for individually fitness-damaging behaviour.

Because BPC does not recognize that adaptations are not necessarily predicted to produce adaptive outcomes in novel environments, it overestimates the degree to which evolutionary theory predicts behaviour that maximizes fitness and/or utility. The psychological mechanisms governing behaviour are conditional decision-rules that respond to specific environmental information by producing specific psychological and behavioural outcomes. Therefore, evolutionary theory casts individuals as "adaptation executers," not "rational choosers" or "fitness maximisers" (Tooby & Cosmides 1992). This framework can, in principle, explain individual performance on a range of decision tasks; it can explain why people are good at reasoning about some problems and not others, why they make particular kinds of systematic mistakes, and so on. However, Gintis regards evolutionary psychology as predicting that individuals are rational actors who choose the available course of action that they expect will maximise their fitness. Therefore, according to Gintis, irrationality presents a problem for evolutionary theory, one that BPC attempts to solve by incorporating a host of ad hoc – albeit well-measured – anomalies, constraints, preferances, and biases.

BPC explains cultural transmission in terms of psychological mechanisms for various forms of imitation (prestige-biased, popularity-related, etc.), and this appears to be an effort to ground cultural evolution in genetic/brain evolution. Although we endorse this effort, the ways in which psychological mechanisms generate and embrace/reject cultural characteristics are far richer than can be captured by an emphasis on general-purpose imitation mechanisms alone. Differences in cultural evolutionary trajectories are largely the products of psychological mechanisms responding to different environments. For example, in an environment offering many benefits from group cooperation (one characterized by large game, coalitional conflict, etc.), psychological adaptations for cooperation may be deployed more often than in an environment offering few such benefits, and the environment offering more benefits may therefore elicit a more cooperative culture. In both cultures, certain specialized imitation mechanisms may indeed help individuals learn how to behave; however, an emphasis on imitation alone would overlook the psychological mechanisms that determined each culture's orientation in the first place. Moreover, a more useful theory of, say, prestige-biased imitation would illuminate not just the potential benefits of imitating successful individuals, but also the potential costs (for example, if a subordinate acts as if he or she has as much power as a dominant individual, this may anger the dominant individual).

BPC's ability to predict sophisticated imitation processes is also limited by its failure to recognize the potential importance of intragenomic conflict in decision-making and social behaviour (Haig 2000). A focus on strategic genes influencing the design of psychologcal mechanisms helps elucidate why imitating a model individual may have differiential costs to matrilineal versus patrilineal inclusive fitness (Brown 2001; Trivers 2000). Indeed, BPC is predictively mute on all forms of intragenomic conflict, and therefore on how individual preferences may conflict and/or be suppressed by rival psychological mechanisms (e.g., see Haig [2003] on intrapersonal reciprocity). This suggests that BPC is not up to the task of uniting the social and natural sciences, especially in the age of genomics.

In conclusion, we favour increased integration among the behavioural sciences. However, BPC would be inhibited in

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

achieving this goal, and in achieving the more ambitious and productive goal of integrating the social and natural sciences, because it does not identify the modern theory of adaptation by natural selection as the core integrating principle. Integration would be better accomplished by the non-zoocentric adaptationist framework that already exists (Darwin 1859; Haig 2003; Hamilton 1964; Tooby & Cosmides 1992; Trivers 1971; 1972; 1974; Williams 1966), and it is not clear that BPC contributes to the progress that this framework continues to make.

Information processing as one key for a unification?

DOI: 10.1017/S0140525X07000829

Michael Schulte-Mecklenbeck

Department of Marketing, Columbia Business School, New York, NY 10027.
research@schulte-mecklenbeck.com
http://www.schulte-mecklenbeck.com

Abstract: The human information-acquisition process is one of the unifying mechanisms of the behavioral sciences. Three examples (from psychology, neuroscience, and political science) demonstrate that through inspection of this process, better understanding and hence more powerful models of human behavior can be built. The target method for this – process tracing – could serve as a central player in this building process of a unified framework.

The unification of different scientific disciplines such as economics, biology, psychology, and political science under the rubric of the "behavioral sciences" can ultimately provide a better understanding of human beings' cognition, behavior, and interactions. Based on Gintis's framework in the target article, questions would be asked differently, and their answers would have a broader impact. Such a unification demands the rethinking of theoretical and methodological issues in each of the affected disciplines. In this commentary I argue that the detailed inspection of the human information-acquisition process in different disciplines helps in building such a framework. In particular, process tracing can serve as a central method in this endeavor.

Process tracing has been primarily studied in the psychology of decision making (Ford et al. 1989) and uses different methods for recording what information is attended to and when that attention occurs and shifts. Thinking aloud, eye tracking, protocol analysis, and information boards are common methods. They rest on the assumption that the recorded information-acquisition steps resemble closely cognitive processes within the human brain. A substantial body of evidence (Harte et al. 1994; Payne et al. 1993; Russo 1978; Schkade & Johnson 1989) has been developed over the last 20 years to support this claim. Here I will highlight three examples from different domains to show how process tracing methods have been used and what benefits arise in comparison to more traditional input-output models.

The first example uses an information board approach to find underlying patterns in information acquisition when simple gambles are used. Brandstätter et al. (2006) suggested a simple descriptive model of people making decisions between two gambles (with two and five outcomes). This method, called the Priority Heuristic (PH), sets the focus (for two-outcome gambles, in decreasing order) on the minimum gain, probability of the minimum gain, and the maximum gain. Using the PH, the authors predicted choices given the use of the heuristic, and prescribed in detail the sequence in which the different items should be accessed. Johnson et al. (under review) compared the process steps of the PH with their observed usage in a process tracing study using the same gambles. It becomes clear when the data of this study are inspected that there are some predictions of PH actually met in the process data; for example, more attention is set to gains in a first reading phase than to probabilities. However, there are several predictions which do not hold when the process level is examined. One of the stronger predictions PH makes is that there should be no transitions between gains and their corresponding probabilities. However, in the Johnson et al. data, this is the most frequent transition found across the different gambles and can be interpreted as integration of the gain-probability pairs into an expected value.

The second example brings us into the domain of neuroscience. Fellows (2006) used an information board approach to identify differences in information-acquisition strategies in a group of participants with damage in the ventromedial frontal lobe (VFL) in comparison with a healthy group (as well as with a frontal lobe–damaged group where the VFL was still intact). Strong differences between the VFL and the control groups were found in terms of which strategy was used to gather information. Generally, a preference for attribute-based search strategies was found. In the VFL group, a different pattern, with dominating search in alternative-based order, resulted. One important detail of this study is that the absolute amount of information and the time taken to come to a decision were the same in both groups – nevertheless, the underlying strategy in information acquisition differed strongly.

The third example demonstrates the usage of process tracing techniques in the political sciences. Redlawsk (2004) examined the information search process of voters in an election experiment. Because of the dynamic structure such an environment has, a modified version of an information board study was used. In this dynamic information board, cell content is updated during the information search process. This means that the participant has to make two decisions – first, which information is of interest and, second, when is the right time to access certain information. Redlawsk compared a static information board with a dynamic one and found a switch from compensatory to non-compensatory strategies with an increase in complexity. Additionally, more information was acquired for the finally chosen candidate in comparison to the rejected one. Both findings will not surprise scientists working with process tracing, because they are well documented in many studies in this field. The lesson from this study is the applicability of the method in a very different domain than is generally used in decision-making studies – the domain of political science and policy building.

The points I want to make with these examples are twofold. First, the three studies show that despite the different perspectives on human behavior, at least some approaches in psychology, neuroscience, and political science use the same methods to gather insights into human information acquisition. However, the adaptation of new methodologies from other areas often takes a long time, and one should be aware that the cited studies (despite their quite recent publication dates) refer to methods that have existed in psychology for more than 20 years.

The second point is that better models of decisions can be built when the input-output level is left and the process actually happening during the information-acquisition phase of a decision is examined. Put simply, process models of human decision making require process data. Using process methods, we can learn when and where the participant sets her focus in her information search through the timing and number of acquisitions of particular information items. As such, we can get closer to underlying processes in the brain when we observe transitions between information items. All of this information would be unavailable if the level of data collection were confined to only the responses of the participants – that is, to their final choices. The wealth of information participants emit when thinking and deciding is valuable, and perhaps critical, in developing unified models in all of the behavioral sciences.

ACKNOWLEDGMENTS

I want to thank the Swiss National Science Foundation (grant PBFR1-110407) for their financial support, as well as Anton Kühberger and Ryan O. Murphy for comments on the draft.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

More obstacles on the road to unification

DOI: 10.1017/S0140525X07000830

Eric Alden Smith

Department of Anthropology, University of Washington, Seattle, WA 98195-3100.

easmith@u.washington.edu

http://faculty.washington.edu/easmith/

Abstract: The synthesis proposed by Gintis is valuable but insufficient. Greater consideration must be given to epistemological diversity within the behavioral sciences, to incorporating historical contingency and institutional constraints on decision-making, and to vigorously testing deductive models of human behavior in real-world contexts.

The ambition of the vision and the scope of knowledge found in the target article put one in mind of the Enlightenment philosophers. Many of my colleagues in the social sciences would view this as hubris, but I applaud Gintis's effort. Indeed, after a long period of divergence and specialization, scholarship on human behavior appears to be undergoing a period of cross-fertilization and synthesis on a scale not seen since the salons of Britain and France two and a half centuries ago. As Gintis argues, evolutionary theory, in both its biological and game-theoretical forms, is playing a crucial role in this exciting process. Gintis and his colleagues (Sam Bowles, Rob Boyd, Ernst Fehr, among others) are leading figures in this reinvigoration of the Enlightenment project, and have been inspirational to me.

However, there are perhaps greater constraints on the unification of the behavioral sciences than acknowledged in the target article. First, there are significant epistemological divides. Hypothetico-deductive methods reign in economics, and among researchers guided by evolutionary biology (e.g., behavioral ecologists, cultural transmission theorists). But elsewhere in the behavioral sciences this approach is seen as limited or even inappropriate; that is certainly the case among the majority of my colleagues in anthropology. The game-theoretical and optimization tools that Gintis sees as the key to unification are alien to or mistrusted by many behavioral scientists, who adhere to a more empiricist and particularist tradition.

Second, in at least the social science end of the behavioral sciences, historical contingency and institutional constraint are primary foci of analysis and causal explanation; yet these appear to play a minor role in the vision of the behavioral sciences found in the target article. To be sure, historical contingency is central to Darwinian understandings of diversity, and path dependence is a well-known feature of game-theoretical dynamics. Hence, there is no fundamental problem with bringing it within the orbit of a synthetic biocultural framework for analysis of human behavioral diversity. Similarly, institutions (and their close kin, norms) can in principle be understood as the accreted outcomes of past individual and collective decisions, which then act as constraints (in the beliefs, preferences, and constraints [BPC] framework) on present and future decisions. But it is one thing to acknowledge these factors and have an in-principle way of incorporating them into the analytical framework; it is quite another to make them the central object of analysis. The gap between these two is, I'm afraid, a major obstacle in the unification of the behavioral sciences.

Finally, the clarity of formal theory is indispensable, and sorely lacking in many areas of the behavioral sciences. But the most beautiful theories are often wrong or of trivial real-world importance; careful empirical research is absolutely essential to building good theory. Here, I do think Gintis's disciplinary background in economics and decision theory limits his synthesis somewhat. In particular, I find the stress on experimental analysis of strategic decision-making too narrow. Experimental findings remain ambiguous guides to understanding "real life" (naturalistic behavior). As noted in the target article,

> humans are tested in the laboratory under conditions differing radically from real life. Although it is important to know how humans choose in such situations, . . . there is certainly no guarantee they will make the same choices in the real-life situation and in the situation analytically generated to represent it. (sect. 4, para. 6)

Exactly; that is why field studies of human behavior in real-world social and environmental settings, guided by the best theories at hand, are essential. In such studies, the ethnographic methods developed by anthropologists, and the observational methods developed by behavioral ecologists, are crucial for testing and refining the general theories discussed in the target article and elsewhere (Winterhalder & Smith 2000).

None of these roadblocks to unification is insurmountable, and I am sure that to one degree or another Gintis recognizes them and has thought about ways of circumventing them. Perhaps the difference is that, as an anthropologist rather than an economist or game theorist, these issues loom larger in my view of the behavioral sciences and the potential for unification of the same. However, I do believe that the agenda sketched by the target article is both coherent and powerful. The obstacles to unification that I have delineated are interrelated, and to a large extent tackling any one of them helps address the others. In particular, to the extent that the theory and methods advocated in the target article generate vigorous and empirically successful explanations of real-world ethnographic and historical phenomena, resistance to unification will wither and deeper cross-disciplinary understanding will flourish.

The psychology of decision making in a unified behavioral science

DOI: 10.1017/S0140525X07000842

Keith E. Stanovich

Department of Human Development and Applied Psychology, University of Toronto, Toronto, Ontario M5S 1V6, Canada.

kstanovich@oise.utoronto.ca

http://leo.oise.utoronto.ca/~kstanovich/index.html

Abstract: The cognitive psychology of judgment and decision making helps to elaborate Gintis's unified view of the behavioral sciences by highlighting the fact that decisions result from multiple systems in the mind. It also adds to the unified view the idea that the potential to self-critique preference structures is a unique feature of human cognition.

Gintis is right that psychology has largely missed the insight that decision making is the central brain function in humans. He is right again that his unification of the behavioral sciences through an ingenious synthesis of concepts from evolutionary theory and game theory could help greatly to drive this insight home. Putting aside the past sins of the discipline though, I think that psychology can add detail to the unified theoretical structure that Gintis lays out with such skill.

The juncture where the unified view needs elaboration is explicitly pointed to by Gintis – it is the juncture in our ecology where the behavioral consequences of the modern world differ from those of the *environment of evolutionary adaptation* (EEA). As Gintis notes,

> the BPC model is based on the premise that choices are consistent, not that choices are highly correlated with welfare. . . . [F]itness cannot be equated with well-being in any creature. Humans, in particular, live in an environment so dramatically different from that in which our preferences evolved that it seems to be miraculous that we are as capable as we are of achieving high levels of individual well-being. . . . [O]ur preference predispositions have not "caught up" with our current environment. (sect. 9.2, para. 7)

A role for psychology in the unified view is that of emphasizing that mismatches between the modern environment and the

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

EEA necessitate a distinction between subpersonal and personal optimization (Stanovich 2004). A behavior that is adaptive in the evolutionary sense is not necessarily instrumentally rational for the organism (Cooper 1989; Skyrms 1996; Stein 1996; Stich 1990). We must be clear when we are talking about fitness maximization at the subpersonal genetic level in the EEA and utility maximization at the personal level in the modern world. In short, our conceptions of rationality must be kept consistent with the entity whose optimization is at issue.

Distinguishing optimization in the EEA from instrumental rationality for a person in a modern environment opens the way for a constructive synthesis of the unified theoretical view of the target article with the research on anomalies and biases in the judgment and decision-making literature of cognitive psychology and behavioral economics (Samuels & Stich 2004; Stanovich 1999; 2004). The processes that generate the biases (shown not just in the laboratory but in real modern life, as well; see Camerer et al. 2004; Dunning et al. 2004; Hilton 2003) may actually be optimal evolutionary adaptations, but they nonetheless might need to be overridden for instrumental rationality to be achieved in the modern world (Kahneman & Frederick 2002; 2005; Stanovich & West 2000).

Of course, talk of one set of cognitive processes being overridden by another highlights the relevance of multiple-process views in cognitive science, including the dual-process theories now enjoying a resurgence in psychology (Evans 2003; Kahneman & Frederick 2002; 2005; Sanfey et al. 2006; Sloman 1996; Stanovich 1999; 2004) – theories that differentiate autonomous (quasi-modular) processing from capacity-demanding analytic processing. Such views capture a phenomenal aspect of human decision making that any unified view must at some point address – that humans in the modern world often feel alienated from their choices. The domains in which this is true are not limited to situations of intertemporal conflict. This alienation, although emotionally discomfiting, is actually a reflection of an aspect of analytic processing that can contribute to human welfare. Analytic processing supports so-called *decoupling operations* – the mental abilities that allow us to mark a belief as a hypothetical state of the world rather than a real one (e.g., Carruthers 2002; Cosmides & Tooby 2000; Dienes & Perner 1999; Evans & Over 2004; Jackendoff 1996). Decoupling abilities prevent our representations of the real world from becoming confused with representations of imaginary situations that we create on a temporary basis in order predict the effects of future actions. Thus, decoupling processes enable one to distance oneself from representations of the world so that they can be reflected upon and potentially improved. Decoupling abilities vary in their recursiveness and complexity. At a certain level of development, decoupling becomes used for so-called meta-representation – thinking about thinking itself (see Dennett 1984; Perner 1991; Whiten 2001). Meta-representation – the representation of one's own representations – is what enables the self-critical stances that are a unique aspect of human cognition. Beliefs about how well we are forming beliefs become possible because of meta-representation, as does the ability to evaluate one's own desires – to desire to desire differently (Frankfurt 1971; Jeffrey 1974; Velleman 1992).

Humans alone (see Povinelli & Bering 2002; Povinelli & Giambrone 2001) appear to be able to represent not only a model of the actual preference structure currently acted upon, but also a model of an idealized preference structure. So a human can say: I would prefer to prefer not to smoke. The second-order preference then becomes a motivational competitor for the first-order preference. The resulting conflict signals what Nozick (1993) terms a lack of rational integration in a preference structure. Such a mismatched first-order/second-order preference structure is one reason why humans are often less rational than bees are, in an axiomatic sense (see Stanovich 2004, pp. 243–47). This is because the struggle to achieve rational integration can destabilize first-order preferences in ways that make humans more prone to the context effects that lead to the violation of the basic axioms of utility theory. The struggle for rational integration is also what contributes to the feeling of alienation that people in the modern world often feel when contemplating the choices that they have made. People seek more than Humean rationality. They seek a so-called *broad rationality* in which the content of beliefs and desires is critiqued and not accepted as given. That critique can conflict with the choice actually made. The conflict then can become a unique motivational force that spurs internal cognitive reform.

Evolutionary psychology, ecological rationality, and the unification of the behavioral sciences

DOI: 10.1017/S0140525X07000854

John Tooby[a] and Leda Cosmides[b]

[a]*Department of Anthropology, Center for Evolutionary Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106-3210l;* [b]*Department of Psychology, Center for Evolutionary Psychology, University of California, Santa Barbara, Santa Barbara, CA 93106.*

tooby@anth.ucsb.edu **cosmides@psych.ucsb.edu**

http://www.psych. ucsb.edu/research/cep/

Abstract: For two decades, the integrated causal model of evolutionary psychology (EP) has constituted an interdisciplinary nucleus around which a single unified theoretical and empirical behavioral science has been crystallizing – while progressively resolving problems (such as defective logical and statistical reasoning) that bedevil Gintis's beliefs, preferences, and constraints (BPC) framework. Although both frameworks are similar, EP is empirically better supported, theoretically richer, and offers deeper unification.

We applaud Gintis's call for the unification of the behavioral sciences within an evolutionary framework and his objections to the parochialism and lack of seriousness that have allowed traditionalists to continue to embrace mutually incompatible models of individual human behavior. Curiously, however, Gintis comments that prior to his proposal the "last serious attempt at developing an analytical framework for the unification of the behavioral sciences was by Parsons and Shils (1951)" (target article, Note 2). Gintis's proposal might be clearer if he had addressed *evolutionary psychology* (EP) as a fully formulated alternative framework (with a well-developed research tradition involving hundreds of scholars). Either Gintis thinks that the EP framework, with its core "integrated causal model" (Tooby & Cosmides 1992), is not "serious," or the name "evolutionary psychology" misleads him into thinking it is only a branch of psychology rather than an encompassing framework for unifying the behavioral sciences (Cosmides et al. 1992; Tooby & Cosmides 1992).

Evolutionary psychology started with the same objections – to the mutual incompatibility of models across the behavioral sciences, and their inconsistency with evolutionary biology. It also started with the same ambition Gintis expresses – the eventual seamless theoretical unification of the behavioral sciences. Gintis says:

> Psychology could be the centerpiece of the human behavioral sciences by providing a general model of decision making for the other behavioral disciplines to use and elaborate for their various purposes. The field fails to hold this position because its core theories do not take the fitness-enhancing character of the human brain, its capacity to make effective decisions in complex environments, as central. (sect. 3, para. 5)

This exact rationale drove the founding of evolutionary psychology decades ago, but such statements sound time-warped in 2007, when countless researchers across every behavioral science

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

subfield both within and beyond psychology take the "the brain as a decision-making organ" and "the fitness-enhancing character of the human brain" as the central starting point for their research.

There is considerable convergence in the two frameworks (on culture, evolutionary game theory, etc.), but it is illuminating to examine where they diverge. For example, EP would consider evolutionary game theory an ultimate – not a proximate – theory. More importantly, EP rests on the recognition that in cause-and-effect terms, it is the information-processing structure of our evolved neurocomputational mechanisms that is actually responsible for determining decisions. This is because selection built neural systems in order to function as computational decision-making devices. Accordingly, computational descriptions of these evolved programs (for exchange, kinship, coalitions, mating) are the genuine building blocks of behavioral science theories, because they specify their input-output relations in a scientific language that (unlike BPC) can track their operations precisely. For example, kin selection theory defines part of the adaptive problem posed by the existence of genetic relatives; but it is the architecture of the human kin detection and motivation system that controls real decision making, not an optimization function (Lieberman et al. 2007).

The design of these programs is *ecologically rational* (Cosmides & Tooby 1994) rather than classically rational either in Gintis's BPC minimalist sense or in widely accepted stronger senses. Classically, decisions are considered irrational when they depart from some normative theory drawn from mathematics, logic, or decision theory (such as choice consistency, the propositional calculus, or probability theory). Departures are indeed ubiquitous (Kahneman et al. 1982). However, these normative theories were designed to have the broadest possible scope of application by stripping them of any contentful assumptions about the world that would limit their generality (e.g., p and q can stand for anything in the propositional calculus).

Natural selection is not inhibited by such motives, however, and would favor building special assumptions, innate content, and domain-specific problem-solving strategies into the proprietary logic of neural devices whenever this increases their power to solve adaptive problems. These special strategies can exploit the long-enduring, evolutionarily recurrent ecological structure of each problem domain by applying procedures special to that domain that are successful within the domain even if problematic beyond it. These decision-making enhancements are achieved at the cost of unleashing a diverse constellation of specialized rationalities whose principles are often irrational by classical normative standards but "better than rational" by selectionist criteria (Cosmides & Tooby 1994).

Research on the Wason task, for example, indicates that humans evolved a specialized logic of exchange that is distinct from "general" logic – and so produces "faulty" choices. Its scope is limited to exchange, and its primitives are not placeholders for any propositions p and q, but rather *rationed benefit* and *requirement*. It uses procedures whose success depends on assumptions that are true for the domain of exchanges, but not outside it. Because of this, it solves reasoning problems involving exchange that the propositional calculus cannot solve. Evidence indicates that this mechanism is evolved, reliably developing, species-typical, neurally dissociable, far better than general reasoning abilities in its domain, and specialized for reasoning about exchange (Cosmides & Tooby 2005). Indeed, economists might be interested in learning that the neural foundation of trade behavior is not general rationality, but rather, rests on an ecologically rational, proprietary logic evolutionarily specialized for this function. (For comparable analyses of the ecological rationality underlying Ellsberg Paradox-like choices, and an evolutionary prospect theory to replace Kahneman and Tversky's [1979] prospect theory, see Rode et al. [1999].)

The Theory of Mind (TOM) mechanism is a specialization that causes humans to interpret behavior in terms of unobservable mental entities – beliefs and desires (Baron-Cohen et al. 1985). We think that the discipline of economics was built out of this seductive framework through its mathematical formalization, without awareness of the extrascientific reasons why its foundational primitives (beliefs, preferences) seem intuitively compelling while being scientifically misleading. Like BPC, TOM does not see the mind's many mechanisms, resists seeing that many computational elements do not fractionate into either "beliefs" or "preferences," and does not recognize that the "knowledge states" inhabiting these heterogeneous subsystems are often mutually inconsistent (Cosmides & Tooby 2000). The BPC framework is a partial, occasionally useful, ultimate theory of selection pressures that our evolved programs partly evolved to conform to. It is distant from any core model of individual behavior that could unify the behavioral sciences. For that, we need the progressively accumulating product of EP: maps of the computational procedures of the programs that constitute our evolved psychological architecture.

Emotions, not just decision-making processes, are critical to an evolutionary model of human behavior

DOI: 10.1017/S0140525X07000866

Glenn E. Weisfeld[a] and Peter LaFreniere[b]

[a]*Department of Psychology, Wayne State University, Detroit, MI 48202;*
[b]*Department of Psychology, University of Maine, Orono, ME 04469.*
ad4297@wayne.edu **peterlaf@maine.edu**

Abstract: An evolutionary model of human behavior should privilege emotions: essential, phylogenetically ancient behaviors that learning and decision making only subserve. Infants and non-mammals lack advanced cognitive powers but still survive. Decision making is only a means to emotional ends, which organize and prioritize behavior. The emotion of pride/shame, or dominance striving, bridges the social and biological sciences via internalization of cultural norms.

We agree wholeheartedly that evolutionary theory must serve as the basis for unifying the behavioral sciences. Other, specifically behavioral, theories apply only to some limited domain of behavior, such as personality, learning, cultural beliefs, or cognition. Another strength of Gintis's model is his emphasis on neural mechanisms. However, when he focuses on decision-making, he commits the very same error of excluding essential categories of behavior.

If we step back and view behavior from an evolutionary standpoint, it becomes apparent that fitness-enhancing behaviors themselves, rather than decision-making or other cognitive processes, are paramount. Ultimately, selection can operate only on the behavioral consequences for the individual organism. All animals must execute some basic, essential behaviors, such as feeding, respiration, excretion, defense, temperature regulation, and reproduction. This is true even of protozoans, which lack learning or cognition. Only mammals possess a cerebral cortex, seat of most behaviors of interest to Gintis.

Decision making in simple (but often very successful) animals is virtually absent. Behavior consists of responding automatically to releasers as they are encountered. Therefore, Gintis's model would not apply to these animals, or to the stereotypic behaviors of more complex organisms, such as primates' reflexes and facial expressions. Yet all these behaviors are already included in a model of behavior that is truly comparative and emphasizes naturally occurring behaviors – an ethological one.

A model of human behavior that does not easily integrate data from other species, risks excluding all the emerging information about our close genetic relationship to other species. It also risks ignoring the adaptive features of bodily systems that interact with the central nervous system, thus perpetuating the mind-body schism.

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Gintis's model also neglects ontogeny. Tellingly, he states that the mainstays of his model, evolution and game theory, cover ultimate and proximate causation. But Tinbergen (1963) also included ontogeny and phylogeny in his four levels of behavioral explanation. Gintis's model does not easily incorporate the behavior of infants and children, who have inchoate cognitive capacities and yet behave successfully enough to survive. Furthermore, the unity of development and its correspondence with phylogenetic adaptation must be addressed. This means being able to describe how an evolved system emerges from precursors and the processes by which it is transformed and reorganized over the life course to meet adaptational needs.

An ethological model gives prominence to behaviors with great phylogenetic stability, namely, motivated behaviors, or *emotions.* These essential, fitness-enhancing behaviors are guided, in complex organisms, by capacities for learning and cognition. But there is no adaptive value in learning or thinking unless it leads to adaptive behavior. First there was motivation, and only later, in some species, did cognition evolve to enhance the efficiency of motivated behaviors. Phylogenetically, the limbic system, which mediates motivation and emotion, preceded the cortex. Even in humans, the limbic system sends more outputs to the cortex than it receives from it. The cortex is often said to be the servant of the hypothalamus (Wilson 1975).

A model of human behavior that revolves around the emotions would provide a framework for incorporating most essential aspects of behavior. Hunger, thirst, sexual feelings, tactile feelings, tasting, smelling, fatigue, drowsiness, anger, fear, pride and shame, love and loneliness, boredom and interest, and humor appreciation – these are universal affects that prompt our essential adaptive behaviors and have deep phylogenetic roots (Panksepp 1998). They can serve as unifying concepts for many disciplines in the sciences and humanities. Such a model would include the internal as well as external elicitors of affect, the overt behavior that the affect prompts, and the emotional expressions and visceral adjustments that accompany many emotions.

Such a model could incorporate age and sex differences in emotional behavior. The emotions change across the life span. For example, infants possess a sucking drive and a desire for rhythmic vestibular stimulation of the sort experienced when being carried. The sex drive appears at puberty. Various emotions may differ quantitatively between men and women. Emotional pathologies such as depression and conduct disorder vary across age and gender.

A model that centered on the emotions could aid us in characterizing individual and cultural differences. Differences in personality and temperament, including many psychopathologies, are essentially differences in the threshold for various affects. Cultures, economies, and political regimes might be described in terms of their success in addressing various emotional needs. Economic models that reduce human behavior to striving for material goods offer a cramped view of human nature; surely we need to incorporate into our models of well-being intangibles such as esthetic and social factors. We need a current, evolutionary model of human needs and tendencies in order to address the normative questions of the social sciences and humanities. Cognitive and economic models are insufficient.

One great strength of Gintis's model is his inclusion of pride and shame, a neglected emotion involving the orbitofrontal cortex, a limbic structure. Recognizing this emotion, ethologists have argued that striving for approval evolved from dominance striving in other primates (e.g., Mazur 2005; Omark et al. 1980; Weisfeld 1997). Unlike animals, however, humans compete for status mainly in non-combative ways, at least after childhood. Each culture, as Gintis says, socializes its children to adopt values that promote fitness in that particular environment. Individuals who fulfill these values gain fitness advantages, often by helping others and earning their trust and reciprocal help (Trivers 1971). But we also "internalize" other sorts of values, such as those concerning what foods to eat and what dangers lurk. The emotion of pride and shame is not sui generis, a superego; it competes for priority just as do other motives. We seek the approval of others and abide by social values in order to maintain our social status, but if hungry enough, we may steal.

Gintis mentions that his model can explain pathological behaviors such as drug addiction, unsafe sex, and unhealthy diet. However, evolutionists have addressed such "diseases of civilization" effectively without recourse to decision-making concepts (e.g., Nesse & Williams 1994/1996). For example, phobias seem to constitute exaggerated fears of objects that were dangerous in prehistory, such as heights and strangers.

Lastly, Gintis's model privileges laboratory research conducted on isolated individuals performing artificial tasks. This research doubtless helps us to imagine behavior that occurred under the prehistoric social conditions that shaped the human genome. However, we can gain more direct insight by studying spontaneous behavior, especially in forager cultures, given that humans evolved as a collectively foraging species arranged in extended families.

We are driven by our emotions, which are guideposts to fitness. We attend to and remember stimuli with emotional significance. We repeat behaviors that are emotionally rewarding, and avoid aversive actions. Our "errors" in reasoning are often systematic and adaptive, such as self-overrating, which apparently helps maintain self-confidence and feelings of deservedness. Rationality would have evolved only insofar as it served these pre-existing emotions and the adaptive behaviors they prompt. We have labeled ourselves *Homo sapiens*, but it is time to disabuse ourselves of the overemphasis on learning and cognition that has plagued the behavioral sciences since the time of Watson, and philosophy since Descartes.

The indeterminacy of the beliefs, preferences, and constraints framework

DOI: 10.1017/S0140525X07000878

Daniel John Zizzo

School of Economics, University of East Anglia, Norwich NR7 4TJ, United Kingdom.

d.zizzo@uea.ac.uk

http://www.uea.ac.uk/~ec601

Abstract: The beliefs, preferences, and constraints framework provides a language that economists, and possibly others, may largely share. However, it has got so many levels of indeterminacy that it is otherwise almost meaningless: when no evidence can ever be a problem for scientific construct Z, then there is a problem for Z, for nothing can also be considered supportive of Z.

Herbert Gintis and I share a similar language. It is (if with different emphasis) the language of an extended, socially grounded, and cognitively limited version of rational choice – the language of game theory and evolution, and that of experimental and neuroscientific evidence. An achievement of the target article is in enabling readers to see how, among the many disagreements, there is also significant cross-talk going on among different behavioural sciences. Of course, the sceptic may reply that the reason that the author and I share a similar language is because, ultimately, we are both economists.

I am less clear about what is the contribution of the target article's beliefs, preferences, and constraints (BPC) framework beyond the broad recognition of common themes and the exposition of specific views on specific points. It is natural for scientific frameworks, as opposed to specific theories, to have degrees of indeterminacy; but in order to be meaningful they still need to put restrictions on what they can explain. Take rational choice in economics. Rational choice by an individual is any choice

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

that maximizes her utility, as defined by a stable and reasonably parsimonious utility function, subject to the budget constraint. This is a framework with degrees of freedom, as it does not specify the specific utility function or model in relation to which rational choice applies. But the qualification that the utility function should be stable and parsimonious ensures that the framework is not empirically empty (Vanberg 2004): otherwise, any anomaly (e.g., framing) could be rationalized away by allowing an extra term in the utility function explaining away the apparent contradiction.

As a second example, take evolutionary psychology (e.g., Barkow et al. 1992). Evolutionary psychology states that the mind is a collection of specialised genetically hardwired mechanisms (modules), impenetrable to environmental influence. According to evolutionary psychologists, such modules reflect the adaptive solution to maximise the reproductive fitness of humans in the Pleistocene. Their methodology boils down to finding adaptive stories for why any trait X may have been fitness-maximising in the Pleistocene, and consider that as a theory of trait X. Although I share Lewontin's (1990) scepticism about the ability to generally test what may be evolutionary just-so-stories, at least in principle this framework does put restrictions: If it were possible to show that in the Pleistocene trait X could not have been evolved as a specialised module, this would falsify evolutionary psychology.

Unfortunately, the BPC framework, as exposed and defended in the target article, combines the indeterminacies of these and other frameworks, and as a result puts virtually no restriction or constraints on what can be explained.

Assume that a researcher finds apparent BPC anomaly Y (e.g., Joe smokes, or engages in aggressive behaviour as the result of being treated procedurally unfairly). A natural BPC explanation is to say that the action is the result of a fitness function maximisation in the present time (e.g., Joe smokes because he feels he has a better chance with the girls if he does). Alternatively, the explanation would be that, although not fitness-maximising at the present time, the action was fitness-maximising in the past: in the smoking example, the stimulants released with smoking must have had adaptive value at some point in the past. Alternatively, one could say that the agent simply maximises her utility function, which is the by-product of a socially plastic brain. Alternatively, one could assume that genetic and social evolution have interacted in one of a number of possible ways. Alternatively, it could all be a matter of framing. Alternatively, it could be performance error, a black box for virtually everything else.

It is not worrisome that there are multiple explanations within a framework; what is worrisome is that the BPC puts basically no restrictions on what is admissible. Many of the alternatives (such as the evolutionary ones) can be fairly indeterminate and untestable, others can be more precise and testable, a few can even be consistent with each other; but in no sense is testing any of them a test of BPC. This is no comfort for finding the BPC framework persuasive: When no evidence can ever be a problem for scientific construct Z, then there is a problem for Z, for nothing can also be considered supportive of Z.

Take the section 9.6 example of the conjunction fallacy, by which subjects consider the likelihood of a combination of two events as greater than each of the events individually. Gintis thinks that this is a problem just of faulty logic, which I believe is incorrect because the fallacy can be reproduced in a purely behavioural context and is behaviourally robust to learning opportunities (Zizzo 2003; 2005). Assuming that I am right, would this be a problem, however, for the BPC model? The answer is no, since, for example, one could simply say that agents are not good at forming beliefs and this may induce performance errors (the sect. 9.7 defence). Or one could say that what agents do represents a heuristic that has evolved as optimal now. Or, if not now, maybe it has evolved as an optimal heuristic sometime in the past. Or one could even blame our evolved, socially plastic brain as not having received enough environmental training to be able to deal with probability compounding seriously. And so on.

A few specific points: (a) Given the above, BPC can be consistent with more than rational choice; but, also, no evidence for or against rational choice has any implication for the status of BPC. (b) There is no reason why BPC should stand or fall on the status of expected utility theory; rational choice itself is compatible with any conventional approach such as generalised expected utilities, including variants of prospect theory (Starmer 2000). (c) However, prospect theory, which the author supports in section 9.2, *is* incompatible with expected utility theory as endorsed elsewhere. (d) As discussed in Zizzo (2002), what the evidence on dopamine shows is support not for a single, unitary value function (such as expected utility), but rather, for either an adaptive learning mechanism (e.g., Schultz et al. 1997) or purely as an attentional mechanism (e.g., Horvitz 2000; Redgrave et al. 1999a; 1999b). (e) Belief formation is indeed an area where research still needs to be done, given the limitations of rational expectations (Menzies & Zizzo 2006).

Author's Response

Unifying the behavioral sciences II

DOI: 10.1017/S0140525X0700088X

Herbert Gintis

Behavioral Sciences, Santa Fe Institute, Santa Fe, NM 87501; Department of Economics, Central European University, Budapest, H-1051 Hungary.

hgintis@comcast.net **http://www-unix.oit.umass.edu/~gintis**

Abstract: My response to commentators includes a suggestion that an additional principle be added to the list presented in the target article: the notion of human society as a complex adaptive system with emergent properties. In addition, I clear up several misunderstandings shared by several commentators, and explore some themes concerning future directions in the unification of the behavioral science.

The target article offered three main points. First, the core theoretical constructs of the various behavioral disciplines include mutually contradictory principles. Second, this situation should not be tolerated by adherents to the scientific method. Third, progress over the past couple of decades has generated the instruments necessary to resolve the interdisciplinary contradictions. I am gratified to note that no commentator disagreed with the first of these assertions, and most agreed with the second, as well. Many agreed with the third point; and some, including **Jones,** elaborated on a theme only implicit in my analysis, that unification could foster more powerful intra-disciplinary explanatory frameworks. Finally, the remarks of several of the commentators, particularly those of **Arnhart**, **Mesoudi & Laland**, and **Smith**, have induced me to add another basic tool to my proposed framework, that of human society as a *complex adaptive system* (see sect. R2).

My proposed framework for unification included three conceptual units: (a) gene-culture coevolution; (b) evolutionary game theory; and (c) the beliefs, preferences, and constraints (BPC) model of decision-making. I will refer to this nexus of concepts (now including the new entrant, complex adaptive systems theory) as "the framework." The comments focusing on my third point fall neatly into

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

four categories: misunderstandings of, suggested additions to, suggested deletions from, and complete alternatives to the framework.

R1. A unifying bridge, not a unified alternative

The most common misunderstanding of my argument is to consider the framework as an alternative core for all the behavioral disciplines. By contrast, I offered my framework as a bridge *linking* the various disciplines, arguing that where two or more disciplines overlap, they must have, but currently do not have, compatible models. Because this overlap covers only a fraction of a discipline's research agenda, the framework leaves much of existing research and many core ideas untouched. For an example of this misunderstanding, **Ainslie** fears that anti-reductivists will reject the BPC model because it ignores the complexity of human consciousness, to which I respond that the BPC model ignores, but does not contradict, the phenomenon of human consciousness, which is likely an *emergent property* of brain function, partially but not completely analytically modeled on the neuronal level (Morowitz 2002). Similarly, **Gow** asserts that unification under the rubric of game theory "would constitute a step backwards" and that "Language processing, memory, problem solving, categorization, and attention are not easily construed as instances of strategic interaction." **Clarke** expresses similar hesitations. However, I stressed that my unification framework would deeply affect those areas where inferences depend on an explicit model of human decision-making and strategic interaction. Many research areas in each discipline would therefore likely be untouched by unification, at least in our current state of knowledge.

Reacting to my statement that "if decision theory and game theory are broadened to encompass other-regarding preferences, they become capable of modeling all aspects of decision making" (target article, Abstract), **Colman** asserts that "game theory ... cannot adequately model all aspects of interactive decision making." I should have said "contributing to modeling" rather than simply "modeling." Decision theory and game theory provide a universal lexicon for the behavioral sciences, and a methodological tool for performing experiments and systematically collecting data on human behavior. These tools are not a substitute for other research tools or methods.

Colman states that one can understand payoff dominance "only by departing radically from standard assumptions of decision theory and game theory." This view, however, is either simply incorrect, or it depends on an idiosyncratic interpretation of what assumptions are "standard." I accept as "standard" only rationality in the thin sense defined in my framework. Using this alone, in Colman's example of a two-player coordination game with a Pareto-dominant equilibrium, there is a simple decision-theoretic argument that would explain why players choose the Pareto-dominant equilibrium: Each player has a probability distribution over the other's choice, and each player places weight $\geq 1/2$ on the Pareto-dominant choice. Why they choose such weights can be explained by their personal understanding of human motivation and psychology.[1]

In a similar vein, **Mesoudi & Laland** rightly assert that there is no reason for the researcher "to limit himself to just a single theoretical technique [game theory] when others – such as population genetic models [...], agent-based simulations [...], stochastic models [...], and phylogenetic methods [...] – may be more suitable in other cases." **Markman** argues similarly that, "the cultural transmission of information takes place via communication processes that do not seem obviously explicable within game theory." To reiterate: I do not propose game theory, or any other part of the framework, as sufficient for carrying out any particular research objective. I claim game theory is a universal lexicon and the overall framework is a bridge among the disciplines.

One final point of possible misunderstanding deserves mention. **Foss** observes that "quantum theory and relativity theory have resisted all attempts at unification in a single theory: they are, in Gintis's terms, incompatible. Yet physics is unified." In fact, physics is not currently unified, and the task of completing unification is thwarted by the lack of observable data concerning the overlap between gravitational and quantum theory. There is no scandal, however, as the problem is at the forefront of research, as opposed to being discreetly hidden from public view – the unfortunate situation in the behavioral sciences.

R2. Human society is a complex adaptive system

Arnhart suggests that a unified framework should incorporate the fact that the "behavioral sciences are historical sciences of emergent complexity that move through a nested hierarchy of three kinds of historical narratives: natural history, cultural history, and biographical history." **Smith** strikes a similar note in observing that the hypothetico-deductive methods of game theory, the BPC model, and even gene-culture coevolutionary theory are "alien to or mistrusted by many behavioral scientists, who adhere to a more empiricist and particularist tradition." Smith notes that "historical contingency and institutional constraint are primary foci of analysis and causal explanation; yet these appear to play a minor role in the vision of the behavioral sciences found in the target article" and goes on to say "the ethnographic methods developed by anthropologists, and the observational methods developed by behavioral ecologists, are crucial for testing and refining the general theories discussed in the target article." **Kennedy**'s remarks are especially apposite in this regard. Cognitive anthropology is quite well suited to interface with gene-culture coevolution and the BPC model, and its special value lies in its ability to model culture and psychology at a level that fills in the black box of physical instantiation of culture in coevolutionary theory. I am particularly attracted to the fragmented concept of culture in this approach, and its recognition that culture must be validated in everyday life or it will be rejected with high probability (Bowles & Gintis 1986; Gintis 1980). **Glassman** adds that my framework "misses ways in which psychology and neuroscience ... may go further beyond gene-culture dualism by articulating how varieties of living systems ... evolve sufficiently as unitary targets of selection to mediate higher-level complex systems."

I believe these valid concerns can be met by characterizing human society as a *complex adaptive system*. A

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

complex system consists of a large population of similar entities (in our case, human individuals) who interact through regularized channels (e.g., networks, markets, social institutions) with significant stochastic elements, without a system of centralized organization and control (i.e., if there is a state, it controls only a small fraction of all social interactions, and itself is a complex system). We say a complex system is "adaptive" if it evolves through some evolutionary (e.g., genetic, cultural, agent-based), process of hereditary reproduction, mutation, and selection (Holland 1975). To characterize a system as "complex adaptive" does not explain its operation, and does not solve any problems. However, it suggests that certain modeling tools are likely to be effective that have little use in a non-complex system. In particular, the traditional mathematical methods of physics and chemistry must be supplemented by other modeling tools, such as agent-based simulation and network theory, as well as the sorts of historical and ethnographic research stressed by **Arnhart, Smith**, and others.

Such novel research tools are needed because a complex adaptive system generally has *emergent properties* that cannot be analytically derived from its component parts. The stunning success of modern physics and chemistry lies in their ability to avoid or strictly limit emergence. Indeed, the experimental method in natural science is to create highly simplified laboratory conditions, under which modeling becomes analytically tractable. Physics is no more effective than economics or biology in analyzing complex real-world phenomena in situ. The various branches of engineering (electrical, chemical, mechanical) are effective because they recreate in everyday life artificially controlled, non-complex, non-adaptive environments in which the discoveries of physics and chemistry can be directly applied. This option is generally not open to most behavioral scientists, who rarely have the opportunity of "engineering" social institutions and cultures.

R3. Emerging transdisciplinary research themes

Several commentators suggest that my proposed framework can foster novel transdisciplinary research themes. **Bentley**, **Glassman**, and **Mesoudi & Laland** outline synthetic models of cultural transmission. **Danielson** offers encouraging observations of how philosophers and behavioral scientists can collaborate in developing ethical theory. **Gow** echoes a theme of **Krueger** in hoping that "rather than dismissing all deviations from the predictions of the model as 'performance errors' [. . .], game theorists could improve their models by addressing how cognitive mechanisms produce systematic variation in performance." **McCain** relates performance error to the beliefs side of the BPC model, saying "Gintis's strategy is to introduce beliefs as an autonomous factor in decisions along with preferences and constraints, and to suggest that well-known empirical anomalies in rational action theory can be isolated as errors in beliefs." This is correct. **Schulte-Mecklenbeck** adds that "The human information-acquisition process is one of the unifying mechanisms of the behavioral sciences," and suggests that "process tracing [. . .] could serve as a central player in this building process of a unified framework." He notes that psychologists are developing descriptive models of choice when the expected utility theorem fails. **Stanovich** elaborates on a similar theme, suggesting that there may be a constructive synthesis of my framework and the anomalies and biases literature. "The processes that generate the biases," he notes, "may actually be optimal evolutionary adaptations, but they nonetheless might need to be overridden for instrumental rationality to be achieved in the modern world." I agree.

Pepper argues persuasively for an increased breadth of sociobiology, covering all orders of biological life. "In biology," he observes,

> key empirical breakthroughs have revealed that what we now recognize as (selfish) individual organisms originated as intensely and elaborately cooperative collectives. . . . Now that we are aware of the enormous potential for ongoing conflict within organisms, both among cells . . . and among genes [. . .], the very existence of organisms that are well-integrated enough to act, selfishly or otherwise, is a testament to the importance of cooperation in both the processes and the outcomes of evolution.

E. O. Wilson (1975) unleashed a furor in suggesting that human sociality is a part of biological sociality (Segerstråle 2001), quite as ferocious as the one unleashed by Darwin when the latter suggested the evolutionary communality of humans and apes (Dennett 1996). The idea of sociobiology is now, however, sufficiently mature to be integrated into the mainstream behavioral sciences.

Stanovich notes that:

> Humans alone . . . appear to be able to represent . . . a model of an idealized preference structure. . . So a human can say: I would prefer to prefer not to smoke. The second-order preference then becomes a motivational competitor for the first-order preference. . . . The conflict then can become a unique motivational force that spurs internal cognitive reform.

I particularly like this research direction, because it was the theme of a chapter of my Ph.D. dissertation some 36 years ago. At the time, however, I could not conceive of how one would study this phenomenon. Now, however, much underbrush has been cleared, and a way forward, integrating economics, psychology, and neuroscience, seems eminently possible.

Weisfeld & LaFreniere argue that

> An evolutionary model of human behavior should privilege emotions: essential, phylogenetically ancient behaviors that learning and decision making only subserve. Infants and non-mammals lack advanced cognitive powers but still survive. Decision making is only a means to emotional ends, which organize and prioritize behavior. The emotion of pride/shame, or dominance striving, bridges the social and biological sciences via internalization of cultural norms. (Weisfeld & LaFreniere commentary Abstract)

We have begun to work on this central transdisciplinary issue (Bowles & Gintis 2004b), but this is a beginning only.

Finally, I would like to endorse **Markman**'s plea that we study how individuals form mental representations, although I do not agree with his view that, "Because questions of mental representation do not fall naturally out of formulations of game theory, research driven by this framework is likely to gloss over issues of representation." Game theorists perhaps have treated "beliefs" and "mental frames" in a rather cavalier manner, but no one has claimed their unimportance for game theory. There will be no shortage of game theorists willing to work in a

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

transdisciplinary setting on this problem (Binmore & Samuelson 2006; Samuelson 2001).

R4. Terminological misunderstandings

Some misunderstanding is based on my using terms in highly specific ways, and others interpreting these terms in different ways. I offer the following in the way of clarification.

I use the term "self-regarding" rather than "self-interested" (and similarly for "other-regarding" and "non-self-interested") for a situation in which the payoffs to other agents are valued by an agent. For instance, if I prefer that another agent receive a gift rather than myself, or if I prefer to punish another individual at a cost to myself, my acts are "other-regarding." I use this term to avoid two confusions. First, if an agent gets pleasure (or avoids the pain of a guilty conscience) from bestowing rewards and punishments on others, his behavior may be rightly termed "self-interested," although his behavior is clearly other-regarding. Second, some behavioral scientists use the term "self-interest," or "enlightened self-interest," to mean "fitness maximizing." By contrast, I generally use terms referring to the behavior of an agent as *proximate* descriptions, having nothing to do with the *ultimate* explanations of how this behavior might have historically come about as a characteristic of the species. For example, one can observe other-regarding behavior in the laboratory, although there are likely evolutionary explanations of why it exists.

Historically, the *inclusive fitness* associated with a behavior has been considered equivalent to the impact of behavior on close relatives, according to the *kin selection* theory of William Hamilton and others, in which relatedness is measured by "common descent" and appreciable levels of relatedness are confined to close relatives (Hamilton 1964). I used the term "inclusive fitness" in this manner in the target article. More recently, several population biologists have recognized the rather complete generality of the inclusive fitness concept, and it is now coming to be used to mean the total effect of a behavior on the population pool of genes that favor this behavior (Fletcher & Zwick 2006; Grafen 2006). In this newer sense, an inclusive fitness of less than unity ensures that a behavior cannot evolve. Either use of the term is acceptable, but if both uses are made, the results are typically undesirable. In this response, I use the term "kin-selection inclusive fitness" to refer to the older usage, and "total-impact inclusive fitness" to refer to the newer.

Finally, an older literature treats gene-level, individual-level, and group-level selection as referring to *the entity that is selected for or against* in an evolutionary dynamic (Mayr 1997). However, any fitness measurement in terms of group-level categories can be reorganized using individual-level categories; and similarly, and individual-level fitness measurement can be written in gene-level terms. Ultimately, unless a behavior increases the expected number (i.e., total-impact inclusive fitness) of genes in the population involved in the behavior, the behavior cannot grow, except possibly in the short term through random diffusion (Kerr & Godfrey-Smith 2002; Wilson & Dugatkin 1997). Therefore, both in the target article and in my remarks below, I use the term "multi-level selection" differently, as follows.

The fitness of a gene depends on the characteristic environment in which it evolves. If the description of the environment relevant to measuring the fitness of a gene does not depend on interactions with other genes, the gene-centered accounting framework is appropriate. Suppose, however, a number of genes act in concert to produce a phenotypic effect. Then, the complex of genes itself is part of the environment under which the fitness of each gene is measured. This complex of genes (which may be localized at the level of the individual) may be best analyzed at a higher level than that of the single gene.

In species that produce complex environments (e.g., beaver dams, bee hives), these environments themselves modulate the fitness of individual genes and gene complexes, so are best analyzed at the level of the social group, as suggested in niche construction theory (Odling-Smee et al. 2003). Gene-culture coevolutionary theory, which applies almost exclusively to our species, is a form of niche construction theory in which cultural rules, more than genetically encoded social interactions, serve to modulate the fitness of various genes and gene complexes. Gene-culture coevolution is therefore group selection, although it must be remembered that the whole analysis of genetic fitness even in this case can be carried out at the level of the individual gene, just so the social context is brought in as relevant to fitness.

In considering group selection in the evolution of human altruism, it is important to distinguish between "hard" and "soft" group selection. The former conforms to the traditional notion of the altruist being disadvantaged as compared with his non-altruist group mates, but altruists as a whole have superior population-level fitness because groups with many altruists do better than groups with few. This form of "hard" (between- versus within-group) selection, exemplified by the use of Price's famous equation (Price 1970), probably is important in the case of humans, especially because human culture reduces the within-group variance of fitness and increases the between-group variance, hence speeding up group-level selection. However, hard group selection is not necessary for my analysis of altruism.

The second, less demanding, form is "soft" group selection, in which altruists are not less fit within the group, but groups with a high fraction of altruists do better than groups with a lower fraction. The forms of altruism that I document in the target article could have evolved by a soft group selection mechanism alone (Gintis 2003b). For instance, suppose social rules in a particular society favor giving gifts to the families of men honored for bravery and killed in battle. Suppose these gifts enhance the survival chances of a man's offspring or enhance their value as mates. The altruism of individuals in this case can spread through weak group selection, leading to more and more human groups following this rule. This is surely group selection, but of course could just as easily be accounted for as individual selection, or even gene selection, as long as the role of social rules in affecting fitness is kept in mind.

R5. Misunderstandings of the BPC model

The BPC model is not an *explanation* of choice behavior, but rather a *compact analytical representation* of behavior. The BPC is, in effect, an analytical apparatus that exploits

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

the properties of choice transitivity to discover a tractable mathematical representation of behavior. A good BPC model is one that predicts well over a variety of parametric conditions concerning the structure of payoffs and the information available to agents. It is often extremely challenging to develop such a model; but when one is discovered, it becomes widely used by many researchers, as in the cases of the expected utility theorem (Mas-Colell et al. 1995; Von Neumann & Morgenstern 1944), prospect theory (Gintis, in press b; Kahneman & Tversky 1979), and quasi-hyperbolic discounting (Laibson 1997; Laibson et al. 2004). The objections that the BPC model is too general, or too specific, or false because people are not rational, are misguided through lack of appreciation of this point.

R5.1. Is the BPC model too general?

Zizzo claims that "The BPC framework . . . puts virtually no restriction or constraints on what can be explained," arguing: "Assume that a researcher finds apparent BPC anomaly Y. . . . A natural BPC explanation is to say that the action is a result of . . . [various alternatives]." However, BPC does not attempt to *explain*, only *represent*. Moreover, it is extremely difficult to discover an adequate representation of some behaviors, and many remain without adequate models, so the notion that the framework can explain virtually anything is just wrong. Similarly, **Hodgson** says, "I do not argue that choice consistency [. . .] is refuted by the evidence. Instead, I uphold it would be difficult in practice to find any evidence strictly to refute this assumption." This is true, because, given transitive preferences, the BPC model is either more or less successful as a tool of scientific discovery. Further on in the commentary, Hodgson continues, "Gintis can point to many examples of its [i.e., the rational actor framework's] apparent success, not only in economics, but also in biology, sociology, and elsewhere. . . . [However, these successes] depend critically on assumptions that are additional to typical axioms of rationality." This is quite correct, but it is no more a critique than to say differential equations are not useful unless we specify the relevant functional forms and parameters. Hodgson asserts that the problem is "explaining where beliefs and preferences come from. . . . From an evolutionary perspective, they can no longer be taken as given." I could not agree more. The BPC is an analytical tool of proximate causality that is useful when there is preference transitivity over some, possibly non-obvious, choice space. It is too general a tool to do any heavy lifting without numerous auxiliary assumptions, and it does not deal with ultimate causality at all. The BPC model shares these properties with all analytical tools, unless we believe in some version of Kant's *synthetic a priori*, which I do not. The evolutionary roots of human behavior have been a primary research area for my colleagues and myself (Bowles & Gintis 2004a; Bowles et al. 2003; Boyd et al. 2003; Gintis 2000d; 2003b; in press b), but the major tools involved include gene-culture coevolution, as well game theory and the BPC model.

Hodgson also complains that "the meaning of 'rationality' is undermined when it is applied to all organisms simply on the basis of the existence of consistent behavior," and that I evacuate "the term 'belief' of much of its meaning when [I suggest] that it applies to all organisms." Accusing me of this is particularly ironic, since I introduced the BPC terminology precisely to avoid the intellectual baggage associated with the term "rational." Moreover, it is plausible to attribute beliefs to a variety of species, although for animals without brains or otherwise lacking the capacity to form mental representations of their life-world, the use of the term is at best a harmless turn of the phrase.

Hodgson's final point is that, compared with the BPC model, there is "an alternative": namely, "What are common to all organisms are not beliefs but *behavioral dispositions*." Of course, unless these dispositions lead to preference intransitivity, they are not an alternative at all. In those (presumably few) cases where preference intransitivity over any plausible choice space fails, I am happy to move to "behavioral dispositions."

R5.2. Is the BPC model too specific?

Zizzo comments that "there is no reason why BPC should stand or fall on the status of expected utility theory; rational choice itself is compatible with any conventional approach such as generalised expected utilities, including variants of prospect theory." I fully agree, and plausible evolutionary arguments can be given for such variants, including prospect theory (Gintis, in press b), although for reasons presented in the target article, I think that evolutionary arguments make expected utility theory the default case.

Bentley objects that "in the beliefs, preferences, and constraints (BPC) model, the assumption that human decisions have an optimal value [. . .] neglects how many behaviors are highly culturally dependent and individually variable." This is not correct. In the target article I show that the assumption of transitivity, plus a few technical conditions, implies the existence of a utility function that represents the agent's choices. Of course, this utility function will be culturally dependent and individually variable.

Continuing in the same vein, **Bentley** asserts that, "Complex choices can be fundamentally different from simple two-choice scenarios. . . . BPC would seem to work best in cases where the complexity of choices is lowest." I agree that the BPC model may not help us predict complex choices in a real-world complex adaptive system. But, neither can anything else. On the other hand, the BPC model suggests general ways such choices might react to parameter shifts. For example, no matter how complex the choice situation, an increase in the cost of taking one option should decrease the probability that that option is taken, so we can obtain a quantitatively accurate elasticity of response to the cost in question. The value of such elasticities for predicting behavior should not be underestimated.

For instance, **Weisfeld & LaFreniere** note that "Gintis mentions that his model can explain pathological behaviors such as drug addiction, unsafe sex, and unhealthy diet. However, evolutionists have addressed such 'diseases of civilization' effectively without recourse to decision-making concepts." I argued that they have addressed these issues *ineffectively* precisely because without the BPC model there is no way to aggregate their effects and predict how behavior will respond to changes in social variables under policy control. I made this point clearly in my discussion of drug addiction in the target

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

article, showing that the BPC model allows us to carry out effectiveness studies of various alternative policies (incarceration, taxation, decriminalization). Evolutionists who reject the BPC model have little to contribute to social policy analysis because *ultimate causality does not reveal proximate causality*.

R5.3. Is human behavior irrational?

Despite my attempt to carefully delimit these terms, confusion still reigns concerning the use of the terms "rational" and "maximize." **Price, Brown, & Curry** [**Price et al.**] argue that individuals are "adaptation executers," whereas the BPC model portrays individuals "as rational actors who choose the available course of action that they expect will maximise their fitness." However, as I made clear in the target article, being a "rational chooser" follows from, and is certainly not incompatible with, being an "adaptation executer." Moreover, the notion that I suggested in the target article (or anywhere else) that individuals make choices that "they expect will maximise their fitness" is a bizarre and outlandish attribution indeed. Acting to maximize fitness does not explain much human behavior.

R5.4. Can the BPC model deal with intergenomic conflict?

Price et al. argue that the BPC model "is predictively mute on all forms of intragenomic conflict, and therefore on how individual preferences may conflict and/or be suppressed by rival psychological mechanisms." They conclude from this that "BPC is not up to the task of uniting the social and natural [*sic*] sciences, especially in the age of genomics." However, the framework I offer does not consist of the beliefs, preferences, and constraints model alone. It includes evolutionary biology in general, and gene-culture coevolution in particular, which allows us not only to deal with intragenomic conflict, but also to model the resulting human behaviors analytically using the BPC apparatus (deQuervain et al. 2004; McClure et al. 2004).

R5.5. Does brain modularity imply non-rational behavior?

Tooby & Cosmides argue that natural selection favors "building special assumptions, innate content, and domain-specific problem-solving strategies into the proprietary logic of neural devices.... These decision-making enhancements ... are often irrational by classical normative standards." Of course, this is one of the central reasons I gave in the target article for abandoning the "classical normative standards" in favor of the principles of consistency on which the beliefs, preferences, and constraints model depends.

R5.6. The BPC model emerges unscathed

In summary, I believe that an idiosyncratic or traditional version of the rational actor model has drawn objections from several commentators here, rather than my version. The version I outlined in the target article appears to have survived attack. Commentators had no trouble offering objections to some version of the rational actor model (generally an idiosyncratic or traditional version), but not to my version. The version I outlined in the target article appears to have survived attack. I am encouraged that the BPC model can remain among the basic analytical tools capable of bridging the various disciplines.

R6. Critique of gene-culture coevolutionary theory

Gene-culture coevolutionary theory holds that (a) there is a cultural dynamic in human society that obeys the same structural equations as genetic evolution; and (b) human culture is a strong environmental influence on genetic evolution, accounting for human prosocial emotions, other-regarding preferences, and principled (such as honest or truthful) behavior. **Brown & Brown** offer their own work on *selective investment theory* (SIT), which is "an example of how other-regarding preferences can be accommodated by a gene-centered account of evolution." They argue that the "fundamental target of selection is the gene, not the group, the species, or even the individual.... [T]he gene-centered view of evolution can and does support other-regarding preferences; there is no need to buy into the less parsimonious and more controversial notion of group selection." In section 4, I offered a second account to explain how a gene-centered account is not different from an individual-centered or a group-centered account. We now know that gene, individual, and group selection are simply alternative accounting frameworks for explaining the same phenomenon (Fletcher & Zwick 2006; Grafen 2006; Kerr & Godfrey-Smith 2002; Wilson & Dugatkin 1997). As soon as Brown & Brown extend other-regarding preferences to "genetically unrelated coalition partners," they are implicitly moving to an accounting framework above the gene level (genes do not have coalition partners). Selective investment theory was developed by Brown & Brown to explain social bonding in long-term relationships, and the willingness of individuals to engage in costly long-term investment in such relationships. The context for such relationships is generally kin and family, where the larger set of cultural institutions can be taken as given, and the assertion of "gene-centered" evolution has at least a semblance of plausibility – although here, as well, "family" includes unrelated mates, at least, and long-term bonds in such cases might fare better in a family-centered accounting framework. At any rate, gene-culture coevolution applies to a much broader set of human behaviors, propensities, and institutions than does selective investment theory.

R7. Critique of strong reciprocity

As an example of the synergistic interaction of the various elements of my framework for unification, I referred to work by myself and colleagues on *strong reciprocity*, which is a predisposition in humans to cooperate with others, and to punish those who violate the norms of cooperation, at personal cost, even when these costs cannot be repaid. No commentary author disputes the existence of strong reciprocity, but several question my evolutionary interpretation of the phenomenon. **Brown & Brown** assert that my colleagues and I

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

"subscribe to views of evolution and other-regarding preferences that are themselves steeped in controversy." This is true, but hardly a critique. Strong reciprocity has been around only since the year 1998 (Fehr & Gächter 1998), was identified as such two years later (Gintis 2000d), and has drawn the attention of the behavioral science community only in the recent past.

Brown & Brown characterize our position as a situation in which "helpers sacrifice *inclusive* fitness for the good of the group" (the commentators' emphasis). This is an incorrect interpretation of our models. Strong reciprocity may involve the sacrifice of *individual* fitness on behalf of the group, but never total-impact *inclusive* fitness, or the behavior could not evolve. Moreover, there are signaling models of strong reciprocity in which strong reciprocity is individually fitness maximizing (Gintis et al. 2001), or are part of an inseparable behavioral program that is individually fitness maximizing (Gintis 2003a; 2003b).

Burgess & Molenaar claim that kin selection and reciprocal altruism are sufficient to explain human other-regarding behavior, but they do not reveal how these forces explain strong reciprocity. "Enlightened self-interests" asserts **Getty,** "are still ultimately self-interests." This is true, but strong reciprocity is not self-regarding behavior at all, although it may maximize inclusive fitness, which is a completely different matter. A parent who sacrifices for its offspring is not exhibiting enlightened self-interest, for example, unless one wants to redefine self-interest to mean anything with which one shares genes, in which case it is redundant – the term "inclusive fitness" will do. Getty explains that "If costly other-regarding preferences have evolved in response to selection, then somehow or another they are ultimately in the constrained, relative self-interests of the individuals who express these traits." This is exactly what I am asserting is *not* the case. The analytical empirical bases of the traditional bias of biologists against multi-level selection in general, and gene-culture coevolution in particular, are being supplanted by the various new approaches described earlier.

Getty goes on to say that "Hagen and Hammerstein (2006) provide a critique of Gintis's interpretation of the seemingly selfless behavior of human subjects in contrived experimental games." However, Hagen and Hammerstein do not claim to provide a "critique." Rather, they entertain alternative interpretations to our results and suggest future research that might resolve these issues. Nor do they claim that experimental games are "contrived." I am pleased to see that Getty agrees with me on one important point, where he quotes me as saying, "A moral sense helps us be reasonable, prosocial, and prudential concerning our long-term interests" and says that this "seems like a sensible hypothesis" to him. The paper that attempts to make this point (Gintis 2003b), however, does *not* conclude that our moral sense is *limited to* defending our "long-term and enlightened self-interests."

Price et al. argue that my interpretation of strong reciprocity as an adaptation explicable through gene-culture coevolutionary theory is incorrect: "[T]he observation of such behaviour is not a sufficient basis on which to conclude that the behaviour evolved for the purpose of producing a fitness-damaging outcome." First, I do not believe, and I did not argue, that behavior evolves for a "purpose." Second, I did not argue that strong reciprocity is fitness-damaging; I argued that it is *other-regarding*, and I located strong reciprocity among the various human brain adaptations that support moral behavior. Third, as I explained above, adaptations cannot be on balance fitness-damaging to the genes that account for the behavior, although they may reduce the fitness of some individuals who carry the adapted genotype.

Price et al., following the Cosmides-Tooby paradigm in evolutionary psychology, are hostile to gene-culture coevolutionary theory and indeed to any model of selection above the level of the gene. Hence, they argue that complex prosocial behaviors such as strong reciprocity cannot be adaptations, but rather are fitness-reducing behaviors due to novel environments. They compare the other-regarding behaviors exhibited in laboratory settings to environmental novelty alone, giving the example of pornography. This is quite a poor example. First, the capacity to be motivated by artificial visual material may well be an adaptation. Second, the argument my colleagues and I present is not a simple just-so story. Rather, we supply careful arguments as to why strong reciprocity is an adaptation, based on our understanding of the organization of social life in Pleistocene hunter-gatherer groups; based on the neuroanatomy of the human prefrontal cortex, the orbitofrontal cortex, and the superior temporal sulcus; and based on our understanding of the physiological basis of human emotions (see Gintis et al. 2005a and the comments of **Weisfeld & LaFreniere** in this issue of BBS).

Noë proposes "more reflection on other forms of cooperation before deciding that phenomena or evolutionary mechanisms require unique explanations. Cooperation can be found in a breathtaking number of forms in a wide range of organisms." I welcome Noë's examples of mutualism, which help flesh out a general sociobiological theory of cooperation. It remains, however, that humans deploy forms of cooperation (strong reciprocity, and even reciprocal altruism) that are not found, or are very rare, in other species. However, it may well be that these have mutualistic explanations, such as costly signaling (Gintis et al. 2001).

R8. The evolutionary psychology critique

We are all evolutionary psychologists, but we do not all subscribe to the particular set of doctrines espoused by **Tooby & Cosmides**. These authors recognize the many communalities between my framework and the ideas they developed in their seminal work. My proposed framework depends critically on their pioneering efforts. However, they claim, as do their colleagues **Price et al.,** that the fruit of their labors are *necessary* and *sufficient* to unify the behavioral sciences. "The EP [evolutionary psychology] framework," they write, is ". . . an encompassing framework for unifying the behavioral sciences." This is not the case.

The claim of universality for evolutionary psychology [EP] flows from the virtually exclusive value its proponents place on "ultimate" as opposed to "proximate" causality, and on the univalent emphasis placed on *adaptation* as an ultimate explanatory mechanism. "Adaptation by natural selection," assert **Price et al.,** "is a necessary and sufficient framework for unifying the social and natural

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

[*sic*] sciences." They do not attempt to justify this assertion, and indeed, I do not believe it can be justified. For one thing, many behavioral disciplines stress proximate causality, and are indifferent to ultimate issues unless these provide fruitful hypotheses for proximate modeling. In short, many behavioral sciences are interested in *how things work*, not *how they got that way*. Evolutionary theory is incapable, even in principle, of supplying answers to such proximate questions. For another, human society is a complex adaptive system with emergent properties and forms of stochasticity that defy explanation in terms of natural selection alone.

The evolutionary psychologists working in the tradition of Cosmides and Tooby reject the BPC model because it is a proximate mechanism. Indeed, they attempt to discredit my framework by *identifying* it with the BPC model, despite the fact that I clearly state that it is one of *several* fundamental unifying principles. The BPC model should not be compared with these authors' adaptationist program for the simple reason that the former deals with proximate and the latter with ultimate causality. **Tooby & Cosmides** claim that evolution created "neurocomputational mechanisms" that actually make decisions, rather than systems of transitive preferences, as favored by the BPC model. They suggest that "computational descriptions of these evolved programs [. . .] are the genuine building blocks of behavioral science theories, because they specify their input-output relations in a scientific language that (unlike BPC) can track their operations precisely." This is incorrect. If payoffs to various decisions are frequency dependent (as they generally are, and as postulated in game theory), then no neural structure can explain how decisions are made without reference to the frequency distribution of other agents. This, the BPC model and game theory can do, whereas evolved computational mechanisms are incapable of doing so by definition – they do not include all the relevant decision variables.

The objection made by **Tooby & Cosmides** and their coworkers to the BPC model is that evolution produces highly modular solutions to particular evolutionary problems, so that the brain becomes a collection of specialized modules, each devoted to a particular evolutionarily relevant task. This is true; but humans are capable of discovering novel solutions to problems never before encountered, so that the brain enjoys a *generalized intelligence* that cannot be captured by the discrete modular theory (Geary 2005). This generalized capacity for solving novel problems allows experimentalists to vary the parameters (constraints, information) of a problem and infer from the subjects' choices the nature of the preference function that summarizes their decision-making structures. The extreme modularity proposed by EP is an impediment to EP serving as a bridge across disciplines.

R9. Past attempts at a unification of the behavioral sciences

My proposed unification project accepts and respects that the various behavioral disciplines define their particular research objectives, and carry them out for the most part, without regard to what occurs in other disciplines. Only where their objects of study *overlap* are the requirements of interdisciplinary consistency currently not met. Generally, this is in the area of human decision-making and strategic interaction. While my concept of unification is limited to providing interdisciplinary consistency, its major value is likely to be the increase in explanatory power of both trans- and intradisciplinary work. Some commentators hold a different conception of unification. **Hammond** holds that we already have unification because of "the behavioral sciences' ironclad commitment to a methodology that prevents valid generalization." I argue that no single methodological commitment is sufficient to unify a set of disciplines that have conflicting models of human behavior. **Colman** offers "the theory of operant conditioning" as an alternative. I cannot conceive of how this principle might resolve conflicts among the disciplines. **Clarke** refers to "the structuralist social theories developed by Althusser, Poulantzas, and others, in the 1960s and 1970s" as candidates. I did not include these thinkers because their model of the individual does not so much solve as sweep under the table the contradictions among models of decision making and strategic interaction by asserting the standard structuralist denial of individual agency. Finally, I described earlier why evolutionary psychology in the Cosmides and Tooby tradition, with its central concern with ultimate explanation, does not address disciplines whose main concern is proximate explanation.

R10. Points of contact

Burgess & Molenaar suggest that I express an "objection to reductionism" because I deny that "behavioral science in any sense *reduces* to biological laws." I am quite in favor of reducing complex phenomena to the interaction of their simpler parts whenever possible. However, there are frequently emergent phenomena in moving from biology to the social sciences that have not been successfully analyzed in purely biological terms, and are unlikely to be so in the foreseeable future (Maynard Smith & Szathmary 1995/1997; Morowitz 2002). I never expressed such an objection, nor do I have one.

Getty asks, "What does it mean to have your research 'informed' by fundamental laws, but not 'reduced' to those laws?" That A informs B but B does not reduce to A means that the explanatory and modeling principles of A are useful in B, but there are properties of the systems studied by B that cannot be explained using concepts from A alone. Getty goes on to remark that I do not "get around to showing how useful the proposed unified theoretical framework could be." My claim is that unification will resolve contradictions across disciplines, and that the resulting models will be much more powerful than the array of heterogeneous, incompatible models. I exhibit this possibility in my treatment of strong reciprocity.

Glassman charges that my discussion of the fitness enhancing decision-making ability of human "complex brains" contains the homunculus fallacy. This is an unwarranted charge. Here is what I said: "*The brain evolved because more complex brains, despite their costs, enhanced the fitness of their bearers*" (my emphasis in the target article). Where is the homunculus buried in this bland statement?

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Markman observes that "the concept of a meme is an interesting metaphor for communication drawn from evolutionary theory, but it is hardly a viable theory of cultural transmission of ideas." This is true, but I did not endorse memetics, which is inconsistent with gene-culture coevolution, which I do endorse.

R11. Conclusion

In my target article, I argued that the core principles of the behavioral disciplines contain falsehoods, and only the narrowest of behavioral scientists can be ignorant of this fact. Yet, this situation is accepted with bland equanimity. "Truth," Spinoza (1677/2005) once noted, "is true in itself; it does not depend on any argument for its truth" (*Ethics* II, Prop. 43, Scholium). I interpret this to mean that having substantial grounds for the truth of a proposition is sufficient to explain why an individual accepts its truth. The obverse, however, is also correct: that individuals hold false or incompatible propositions to be true does require explanation.

Why behavioral scientists do not object to the unfounded, and indeed implausible, beliefs of their counterparts in other disciplines is a problem of a different order. I have not attempted to explain this fact here or elsewhere, but it deserves some passing mention, as unification may involve attacking and overturning the social and psychological bases of what might be termed "malignant tolerance." Tolerance is benign, indeed admirable, when it promotes cultural and religious diversity. In science, also, openness to new and even extravagant ideas is desirable. But routinized tolerance of incompatible scientific propositions is highly malignant. For it necessarily replaces the search for truth with an unhealthy but professionally safe splintering of this search into disciplinary fiefdoms where camaraderie among the like-minded reigns supreme, preaching to the choir is de rigueur, and outsiders attempting to impose external standards are deeply resented.

Why do biologists use agent-based models, whereas economists consider them unworthy of recognition? Why do sociologists ignore the brilliant cultural models developed in biology? Why do some disciplines tolerate learning-by-experience models but not imitation models? Why do psychologists delight in interpreting their findings as undermining economic theory? Why do some disciplines barely tolerate analytical methods, while others barely tolerate applied, historical, and ethnographic methods? I do not presume to have answers for these questions.

I suspect that overcoming this lamentable state of affairs will require coordinate action on several fronts. Most important will be concrete transdisciplinary findings that enrich multiple disciplines and subvert disciplinary isolation. It is here that a framework for unification will most surely show its value. Second, established leaders in each discipline must have the scientific integrity to suppress their instinctive support for the very institutions that led them to prominence, by promoting transdisciplinary principles that they perhaps understand only well enough to recognize their power. Third, funding agencies such as the National Science Foundation and the National Institutes of Mental Health must develop a long-range plan for abandoning their ecumenical support for research based on incompatible models in distinct disciplines, and embrace transdisciplinary research that brings the discrepancies among these models center stage.

ACKNOWLEDGMENTS

I would like to thank Samuel Bowles and Eric Alden Smith for comments and the John D. and Catherine T. MacArthur Foundation for financial support.

NOTE

1. In a personal communication, **Colman** refers to Bacharach (1987) in support of his argument. This paper axiomatizes game theory and proves several elegant theorems, including the Nash equilibrium solution concept. However, Bacharach assumes extremely strong rationality axioms going far beyond the "thin" conception of rationality (transitivity) defended in this paper. These include the principle that if p is a theorem, every agent knows that p is a theorem (M2, p. 22). This recursively ensures that each agent knows the others' priors, which is far from empirically acceptable.

References

Letters "a" and "r" appearing before author's initials refer to target article and response, respectively.

Abbott, R. J., James, J. K., Milne, R. I. & Gillies, A. C. M. (2003) Plant introductions, hybridization and gene flow. *Philosophical Transactions of the Royal Society of London B* 358:1123–32. [aHG]

Abel, P. (2003) On the prospects for a unified social science: Economics and sociology. *Socio-Economic Review* 1:1–26. [PD]

Ahlbrecht, M. & Weber, M. (1995) Hyperbolic discounting models in prescriptive theory of intertemporal choice. *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 115:535–68. [aHG]

Ainslie, G. (1975) Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin* 82:463–96. [aHG]

(2001) *Breakdown of will*. Cambridge University Press. [GA]

(2005) Précis of *Breakdown of will*. *Behavioral and Brain Sciences* 28(5): 635–73. [GA]

(2006) What good are facts? The "drug" value of money as an exemplar of all non-instrumental value. *Behavioral and Brain Sciences* 29(2):176–77. [GA]

Ainslie, G. & Haslam, N. (1992) Hyperbolic discounting. In: *Choice over time*, ed. G. Loewenstein & J. Elster, pp. 57–92. Russell Sage. [aHG]

Akerlof, G. A. (1991) Procrastination and obedience. *American Economic Review* 81(2):1–19. [aHG]

Alcock, J. (1993) *Animal behavior: An evolutionary approach*. Sinauer. [aHG]

(2001) *Animal behavior*, 7th edition. Sinauer. [RMB]

Alexander, R. D. (1979) Natural selection and social exchange. In: *Social exchange in developing relationships*, ed. R. L. Burgess & T. L. Huston, pp. 197–221. Academic Press. [RLB]

(1987) *The biology of moral systems: Foundations of human behavior.* Aldine de Gruyter. [aHG, ODJ]

Allais, M. (1953) Le comportement de l'homme rationnel devant le risque, critique des postulats et axiomes de l'école Américaine. *Econometrica* 21:503–46. [aHG]

Allman, J., Hakeem, A. & Watson, K. (2002) Two phylogenetic specializations in the human brain. *Neuroscientist* 8:335–46. [aHG]

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C. & Qin, Y. (2004) An integrated theory of the mind. *Psychological Review* 111:1036–60. [RAM]

Andreoni, J. (1995) Cooperation in public goods experiments: Kindness or confusion. *American Economic Review* 85(4):891–904. [aHG]

Andreoni, J. & Miller, J. H. (2002) Giving according to GARP: An experimental test of the consistency of preferences for altruism. *Econometrica* 70(2):737–53. [aHG]

Anscombe, F. & Aumann, R. (1963) A definition of subjective probability. *Annals of Mathematical Statistics* 34:199–205. [aHG]

Arnhart, L. (1998) *Darwinian natural right: The biological ethics of human nature*. The State University of New York Press. [LA]

(2005) *Darwinian conservatism*. Imprint Academic. [LA]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Arrow, K. J. & Debreu, G. (1954) Existence of an equilibrium for a competitive economy. *Econometrica* 22(3):265–90. [aHG]
Arrow, K. J. & Hahn, F. (1971) *General competitive analysis*. Holden-Day. [aHG]
Atran, S. (2001) The trouble with memes: Inference versus imitation in cultural creation. *Human Nature* 12(4):351–81. [ABM]
(2004) *In gods we trust*. Oxford University Press. [aHG]
Aumann, R. & Brandenburger. A. (1995) Epistemic conditions for Nash Equilibrium. *Econometrica* 65(5):1161–80. [aHG]
Aunger, R. (2004) *Reflexive ethnographic science*. AltaMira. [AM]
Axelrod, R. (1997a) *The complexity of cooperation*. Princeton University Press. [RAB]
(1997b) The dissemination of culture: A model with local convergence and global polarization. *Journal of Conflict Resolution* 41:203–26. [AM]
(1981) *The evolution of cooperation*. Basic Books. [TG]
Axelrod, R. & Hamilton, W. D. (1981) The evolution of cooperation. *Science* 211:1390–96. [aHG]
Bacharach, M. (1987) A theory of rational decision in games. *Erkentniss* 27(1):17–56. [rHG]
(1999) Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics* 53:117–47. [AMC]
(2006) *Beyond individual choice: Teams and frames in game theory*, ed. N. Gold & R. Sugden. Princeton University Press. [AMC]
Ball, P. (2004) *Critical mass: How one thing leads to another*. Heinemann. [RAB]
Bandura, A. (1977) *Social learning theory*. Prentice-Hall. [aHG]
Bargh, J. A. & Chartrand, T. L. (1999) The unbearable automaticity of being. *American Psychologist* 54:462–79. [JIK]
Barkow, J. H., Cosmides, L. & Tooby, J. (1992) *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press. [RLB, DJZ]
Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985) Does the autistic child have a "theory of mind"? *Cognition* 21:37–46. [JT]
Batson, C. D. (1997) Self-other merging and the empathy-altruism hypothesis: Reply to Neuberg et al. (1997). *Journal of Personality and Social Psychology* 73:517–22. [RMB]
(2006) SIT or STAND? *Psychological Inquiry* 17:30–35. [RMB]
Becker, G. S. (1981) *A treatise on the family*. Harvard University Press. [GMH]
Becker, G. S., Grossman, M. & Murphy, K. M. (1994) An empirical analysis of cigarette addiction. *American Economic Review* 84(3):396–418. [aHG]
Becker, G. S. & Murphy, K. M. (1988) A theory of rational addiction. *Journal of Political Economy* 96(4):675–700. [aHG]
Becker, G. S. & Stigler, G. J. (1977) De Gustibus Non Est Disputandum. *American Economic Review* 67(2):76–90. [aHG]
Beer, J. S., Heerey, E. A., Keltner, D., Skabini, D. & Knight, R. T. (2003) The regulatory function of self-conscious emotion: Insights from patients with orbitofrontal damage. *Journal of Personality and Social Psychology* 65:594–604. [aHG]
Bell, D. E. (1982) Regret in decision making under uncertainty. *Operations Research* 30:961–81. [aHG]
Benabou, R. & Tirole, J. (2002) Self-confidence and personal motivation. *Quarterly Journal of Economics* 117(3):871–915. [aHG]
Benedict, R. (1934) *Patterns of culture*. Houghton Mifflin. [aHG]
Bentley, R. A., Hahn, M. W. & Shennan, S. J. (2004) Random drift and culture change. *Proceedings of the Royal Society B* 271:1443–50. [RAB, AM]
Bentley, R.A. & Shennan, S. J. (2003) Cultural evolution and stochastic network growth. *American Antiquity* 68:459–85. [RAB]
(2005) Random copying and cultural evolution. *Science* 309:877–79. [RAB]
Berg, J. E., Dickhaut, J. W & Rietz, T. A (2005) Preference reversals: The impact of truth-revealing incentives. Working Paper, College of Business, University of Iowa. [aHG]
Bernheim, B. D. (1984) Rationalizable strategic behavior. *Econometrica* 52(4):1007–28. [aHG]
Berscheid, E. (2006) SIT: An exercise in theoretical-multitasking. *Psychological Inquiry* 17:35–38. [RMB]
Bhaskar, R. (1979) *The possibility of naturalism: A philosophical critique of the contemporary human sciences*. Harvester. [SC]
Bicchieri, C. (2006) *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press. [PD]
Bikhchandani, S., Hirshleifer, D. & Welsh, I. (1992) A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy* 100:992–1026. [aHG]
Binmore, K. (1987) Modelling rational players: I. *Economics and Philosophy* 3: 179–214. [aHG]
(2006) Why do people cooperate? *Politics, Philosophy and Economics* 5(1): 81–96. [PD]
Binmore, K. G. & Samuelson, L. (2006) The evolution of focal points. *Games and Economic Behavior* 55(1):21–42. [rHG]
Black, F. & Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–54. [aHG]
Blalock, H. M. (1984) *Basic dilemmas in the social sciences*. Sage. [LA]
Blaug, M. (1992) *The methodology of economics: Or, how economists explain*, 2nd edition. Cambridge University Press. [GMH]
Blount, S. (1995) When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes* 63(2):131–44. [aHG]
Bodner, R. & Prelec, D. (2002) Self-signaling and diagnostic utility in everyday decision making. In: *Collected essays in psychology and economics*, ed. I. Brocas & J. D. Carillo, pp. 105–23. Oxford University Press. [aHG]
Boehm, C. (2000) *Hierarchy in the forest: The evolution of egalitarian behavior*. Harvard University Press. [aHG]
Boles, T. L., Croson, R. T. A. & Murnighan, J. K. (2000) Deception and retribution in repeated ultimatum bargaining. *Organizational Behavior and Human Decision Processes* 83(2):235–59. [aHG]
Bonner, J. T. (1984) *The evolution of culture in animals*. Princeton University Press. [aHG]
Borgerhoff Mulder, M. (1998) The demographic transition: Are we any closer to an evolutionary explanation? *Trends in Ecology and Evolution* 13(7):266–70. [aHG]
Borsboom, D. & Dolan, C. V. (2006) Why *g* is not an adaptation: A comment on Kanazawa (2004). *Psychological Review* 113:433–37. [RLB]
Bowles, S. & Gintis, H. (1986) *Democracy and capitalism: Property, community, and the contradictions of modern social thought*. Basic Books. [rHG]
(2000) Walrasian economics in retrospect. *Quarterly Journal of Economics* 115(4): 1411–39. [aHG]
(2004a) The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology* 65:17–28. [rHG, RN]
(2004b) The origins of human cooperation. In: *Genetic and cultural origins of cooperation*, ed. P. Hammerstein, pp. 429–43. MIT Press. [rHG, RN]
(2005a) Prosocial emotions. In: *The economy as an evolving complex system III*, ed. L. E. Blume & S. N. Durlauf. Santa Fe Institute. [aHG]
(2005b) Social capital, moral sentiments, and community governance. In: *Moral sentiments and material interests: The foundations of cooperation in economic life*, ed. H. Gintis, S. Bowles, R. Boyd & E. Fehr, pp. 379–98. MIT Press. [RBG]
Bowles, S., Jung-kyoo, C. & Hopfensitz, A. (2003) The co-evolution of individual behaviors and social institutions. *Journal of Theoretical Biology* 223: 135–47. [rHG]
Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. (2003) Evolution of altruistic punishment. *Proceedings of the National Academy of Sciences USA* 100(6):3531–35. [rHG]
Boyd, R. & Richerson, P. J. (1985) *Culture and the evolutionary process*. University of Chicago Press. [RLB, aHG, AM, RAB]
(1988) The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology* 132:337–56. [aHG]
(1992) Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethology and Sociobiology* 113:171–95. [aHG]
Boyer, P. (2001) *Religion explained: The human instincts that fashion gods, spirits and ancestors*. Heinemann. [aHG]
Brandstätter, E., Gigerenzer, G. & Hertwig, R. (2006) The priority heuristic: Making choices without trade-offs. *Psychological Review* 113(2): 409–32. [MS-M]
Brigden, L. W. & De Beyer, J. (2003) *Tobacco control policy: Stories from around the world*. World Bank. [aHG]
Brown, D. E. (1991) *Human universals*. McGraw-Hill. [LA, aHG]
Brown, J. H. & Lomolino, M. V. (1998) *Biogeography*. Sinauer. [aHG]
Brown, R. M. & Brown, S. L. (2006) SIT stands and delivers: A reply to the commentaries. *Psychological Inquiry* 17:60–74. [RMB]
Brown, S. L. & Brown, R. M. (2006) Selective investment theory: Recasting the functional significance of close relationships. *Psychological Inquiry* 17:1–29. [RMB]
Brown, W. M. (2001) Genomic imprinting and the cognitive architecture mediating human culture. *Journal of Cognition and Culture* 1:251–58. [MEP]
Brumann, C. (1999) Writing for culture: Why a successful concept should not be discarded. *Current Anthropology* 40(Special Supplement):S1–S27. [DPK]
Brunswik, E. (1956) *Perception and the representative design of experiments*. University of California Press. [KRH]
Bshary, R. & Bronstein, J. L. (2004) Game structures in mutualistic interactions: What can the evidence tell us about the kind of models we need? *Advances in the Study of Behavior* 34:59–101. [RN]
Burgess, R. L. (2005) Evolutionary theory and human development. In: *Evolutionary perspectives on human development*, ed. R. L. Burgess & K. MacDonald. Sage. [RLB]
Burgess, R. L. & MacDonald, K. (2005) *Evolutionary perspectives on human development*. Sage. [RLB]
Burks, S. V., Carpenter, J. P. & Verhoogen, E. (2003) Playing both roles in the trust game. *Journal of Economic Behavior and Organization* 51:195–216. [aHG]
Burnham, T. C. & Johnson, D. P. (2005) The biological and evolutionary logic of human cooperation. *Analyse and Kritik* 27:113–35. [RMB]
Burt, A. & Trivers, R. (2006) *Genes in conflict: The biology of selfish genetic elements*. Harvard University Press. [JWP]
Buss, L. (1987) *The evolution of individuality*. Princeton University Press. [JWP]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Camerer, C. (2003) *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press. [aHG]
Camerer, C., Loewenstein, G. & Rabin, M., eds. (2004) *Advances in behavioral economics*. Princeton University Press. [KES]
Camille, N. (2004) The involvement of the orbitofrontal cortex in the experience of regret. *Science* 304:1167–70. [aHG]
Campbell, D. T. (1974) Downward causation in hierarchically organized biological systems. In: *Studies in the philosophy of biology and related problems*, ed. F. Ayala & T. Dobzhansky. University of California Press. [RBG]
Carruthers, P. (2002) The cognitive functions of language. *Behavioral and Brain Sciences* 25:657–726. [KES]
Cartwright, N. (1999) *The dappled world*. Cambridge University Press. [SC]
Casajus, A. (2001) *Focal points in framed games: Breaking the symmetry.* Springer-Verlag. [AMC]
Cavalli-Sforza, L. L. & Feldman, M. W. (1981) *Cultural transmission and evolution.* Princeton University Press. [aHG, AM]
(1982) Theory and observation in cultural transmission. *Science* 218:19–27. [aHG]
Charness, G. & Dufwenberg, M. (2004) Promises and partnership. Working Paper, University of California at Santa Barbara, October 2004. [aHG]
Chen, H.-C., Friedman, J. W. & Thisse, J.-F. (1997) Boundedly rational Nash equilibrium: A probablistic choice approach. *Games and Economic Behavior* 18:32–54. [RAM]
Cheng, P. W. & Holyoak, K. J. (1985) Pragmatic reasoning schemas. *Cognitive Psychology* 17:391–416. [aHG]
Cialdini, R. B. (2001) *Influence: Science and practice*. Allyn & Bacon. [RBG]
Cohen, L. J. (1981) Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences* 4:317–31. [aHG]
Coleman, J. S. (1990) *Foundations of social theory.* Belknap. [aHG]
Colman, A. M. (2003a) Beyond rationality: Rigor without mortis in game theory. *Behavioral and Brain Sciences* 26:180–98. [AMC]
(2003b) Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences* 26(2):139–53. [AMC]
Colman, A. M. & Bacharach, M. (1997) Payoff dominance and the Stackelberg heuristic. *Theory and Decision* 43:1–19. [AMC]
Colman, A. M. & Stirk, J. A. (1998) Stackelberg reasoning in mixed-motive games: An experimental investigation. *Journal of Economic Psychology* 19:279–93. [AMC]
Conlisk, J. (1988) Optimization cost. *Journal of Economic Behavior and Organization* 9:213–28. [aHG]
Conte, R., Hegselmann, R. & Terna, P., eds. (1997) *Simulating social phenomena*. Springer. [RAB]
Cooper, R., DeJong, D., Forsythe, R. & Ross, T. (1990) Selection criteria in coordination games: Some experimental results. *American Economic Review* 80:218–33. [AMC]
Cooper, W. S. (1987) Decision theory as a branch of evolutionary theory. *Psychological Review* 4:395–411. [aHG]
(1989) How evolutionary biology challenges the classical theory of rational choice. *Biology and Philosophy* 4:457–81. [KES]
Cosmides, L. (1989) The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason Selection Task. *Cognition* 31:187–276. [aHG]
Cosmides, L. & Tooby, J. (1994) Better than rational: Evolutionary psychology and the invisible hand. *American Economic Review* 84(2):327–32. [JT]
(2000) Consider the source: The evolution of adaptations for decoupling and metarepresentation. In: *Metarepresentations: A multidisciplinary perspective*, ed D. Sperber, pp. 53–115. [Vancouver Studies in Cognitive Science.] Oxford University Press. [KES, JT]
(2005) Neurocognitive adaptations designed for social exchange. In: *The handbook of evolutionary psychology*, ed. D. M. Buss, pp. 584–627. Wiley. [JT]
Cosmides, L., Tooby, J. & Barkow, J. (1992) Evolutionary psychology and conceptual integration. In: *The adapted mind: Evolutionary psychology and the generation of culture*, ed. J. Barkow, L. Cosmides & J. Tooby. Oxford University Press. [JT]
Crawford, V. P. & Haller, H. (1990) Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica* 58:571–95. [AMC]
Cronk, L. (1999) *That complex whole*. Westview Press. [RAB]
D'Andrade, R. (1995) *The development of cognitive anthropology*. Cambridge University Press. [DPK]
Darley, J. M. & Latané, B. (1968) Group inhibition of bystander intervention in emergencies. *Journal of Personality and Social Psychology* 10:215–21. [JIK]
Darwin, C. (1859) *On the origin of species by means of natural selection*. Murray. [JWP, MEP]
Darwin (1859/1966) *On the origin of species: A facsimile of the first Edition (1966)*, Harvard University Press. [JWP]
(1871) *The descent of man*. J. Murray. [LA]
(1872) *The origin of species by means of natural selection*, 6th edition. John Murray. [aHG]
Dawkins, R. (1976) *The selfish gene*. Oxford University Press. [PD, ABM aHG]
(1982) *The extended phenotype: The gene as the unit of selection.* Freeman. [aHG]
de Waal, F. B. M. (1996) *Good natured: The origins of right and wrong in humans and other animals*. Harvard University Press. [RMB]
Denison, R. F. (2000) Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *American Naturalist* 156:567–76. [RN]
Dennett, D. (1996) *Darwin's dangerous idea.* Simon & Schuster. [rHG]
Dennett, D. C. (1984) *Elbow room: The varieties of free will worth wanting.* MIT Press. [KES]
deQuervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. & Fehr, E. (2004) The neural basis of altruistic punishment. *Science* 305:1254–58. [rHG]
Diekmann, A. (1985) Volunteer's dilemma. *Journal of Conflict Resolution* 29:605–10. [JIK]
Dienes, Z. & Perner, J. (1999) A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences* 22:735–808. [KES]
DiMaggio, P. (1994) Culture and economy. In: *The handbook of economic sociology*, ed. N. Smelser & R. Swedberg, pp. 27-57. Princeton University Press. [aHG]
(1997) Culture and cognition. *Annual Review of Sociology* 23:263–87. [DPK]
Doris, J. M. & Stich, S. P. (2005) As a matter of fact: Empirical perspectives on ethics. In: *The Oxford handbook of contemporary analytic philosophy*, ed. F. Jackson & M. Smith. Oxford University Press. [PD]
Dorris, M. C. & Glimcher, P. W. (2004) Activity in posterior parietal cortex is correlated with the subjective desirability of an action. *Neuron* 44:365–78. [aHG]
Dragoi, V. & Staddon, J. E. (1999) The dynamics of operant conditioning. *Psychological Review* 106:20–61. [AMC]
Dressler, W. W. & Bindon, J. R. (2000) The health consequences of cultural consonance: Cultural dimensions of lifestyle, social support, and arterial blood pressure in an African American community. *American Anthropologist* 102(2):244–60. [DPK]
Dunning, D., Heath, C. & Suls, J. M. (2004) Why people fail to recognize their own incompetence. *Psychological Science in the Public Interest* 5:69–106. [KES]
Durham, W. H. (1991) *Coevolution: Genes, culture, and human diversity.* Stanford University Press. [aHG]
Durkheim, E. (1951) *Suicide, a study in sociology.* Free Press. [aHG]
Eagly, A. & Wood, W. (2003) The origins of sex differences in human behavior. In: *Evolution, gender, and rape*, ed. C. B. Travis. MIT Press. [RMB]
Edgeworth, F. Y. (1881/1967) *Mathematical psychics: An essay on the application of mathematics to the moral sciences.* Augustus M. Kelley. (Original work published in 1881). [AMC]
Eerkens, J. W. & Lipo, C. P. (2005) Cultural transmission, copying errors, and the generation of variation in material culture and the archaeological record. *Journal of Anthropological Archaeology* 24:316–34. [RAB]
Ehrlich, P. (2000) *Human natures: Genes, cultures, and the human prospect*. Island Press. [AM]
Eigen, M. & Schuster, P. (1978) *The hypercycle, a principle of natural self-organization*. Springer-Verlag. [JWP]
Ellsberg, D. (1961) Risk, ambiguity, and the savage axioms. *Quarterly Journal of Economics* 75:643–49. [aHG]
Elster, J. (1979) *Ulysses and the sirens: Studies in rationality and irrationality.* Cambridge University Press. [aHG]
Epstein, J. M. & Axtell, R. (1996) *Growing artificial societies: Social science from the bottom up.* MIT Press. [AM]
Erev, I. & Barron, G. (2005) On adaptation, maximization, and reinforcement learning among cognitive strategies. *Psychological Review* 112:912–31. [RLB]
Eshel, I. & Feldman, M. W. (1984) Initial increase of new mutants and some continuity properties of ESS in two locus systems. *American Naturalist* 124:631–40. [aHG]
Eshel, I., Feldman, M. W. & Bergman, A. (1998) Long-term evolution, short-term evolution, and population genetic theory. *Journal of Theoretical Biology* 191:391–96. [aHG]
Etzioni, A. (1985) Opening the preferences: A socio-economic research agenda. *Journal of Behavioral Economics* 14:183–205. [aHG]
Evans, J. St. B. T. (2003) In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences* 7:454–59. [KES]
Evans, St. J. B. T. & Over, D. E. (2004) *If*. Oxford University Press. [KES]
Fehr, E. & Gächter, S. (1998) How effective are trust- and reciprocity-based incentives? In: *Economics, values and organizations*, ed. L. Putterman & A. Ben-Ner, pp. 337–63. Cambridge University Press. [rHG]
(2000) Cooperation and punishment. *American Economic Review* 90(4):980–94. [aHG]
(2002) Altruistic punishment in humans. *Nature* 415:137–40. [aHG]
Fehr, E., Gächter, S. & Kirchsteiger, G. (1997) Reciprocity as a contract enforcement device: Experimental evidence. *Econometrica* 65(4):833–60. [aHG]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Fehr, E., Kirchsteiger, G. & Riedl, A. (1998) Gift exchange and reciprocity in competitive experimental markets. *European Economic Review* 42(1):1–34. [aHG]
Feldman, M. W. & Zhivotovsky, L. A. (1992) Gene-culture coevolution: Toward a general theory of vertical transmission. *Proceedings of the National Academy of Sciences USA* 89:11935–38. [aHG]
Fellows, L. K. (2006) Deciding how to decide: Ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. *Brain* 129:944–52. [MS-M]
Ferster, C. B. & Skinner, B. F. (1957) *Schedules of reinforcement.* Appleton-Century-Crofts. [AMC]
Fletcher, J. A. & Zwick, M. (2006) Unifying the theories of inclusive fitness and reciprocal altruism. *The American Naturalist* 168(2):252–62. [rHG]
Fishburn, P. C. & Rubinstein, A. (1982) Time preference. *Econometrica* 23(3):667–94. [aHG]
Fisher, R. A. (1930) *The genetical theory of natural selection.* Clarendon Press. [aHG]
Flinn, M. V. (2005) Culture and developmental plasticity: The evolution of the human brain. In: *Evolutionary perspectives on human development*, ed. R. L. Burgess & K. MacDonald. Sage. [RLB]
Ford, J., Schmitt, N., Schechtmann, S., Hults, B. & Doherty, M. (1989) Process tracing methods: Contributions, problems, and neglected research questions. *Organizational Behavior and Human Decision Processes* 43:75–117. [MS-M]
Foss, J. E. (1995) Materialism, reduction, replacement, and the place of consciousness in science. *The Journal of Philosophy* 92:401–29. [JF]
(1998) The logical and sociological structure of science. *Protosociology* 12:66–77. [JF]
Frank, R. H., Gilovich, T. & Dennis, R. (1996) Do economists make bad citizens? *Journal of Economic Perspectives* 10:187–92. [GA]
Frankfurt, H. (1971) Freedom of the will and the concept of a person. *Journal of Philosophy* 68:5–20. [KES]
Fudenberg, D. & Maskin, E. (1986) The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54(3):533–54. [aHG]
Fudenberg, D., Levine, D. K. & Maskin, E. (1994) The folk theorem with imperfect public information. *Econometrica* 62:997–1039. [aHG]
Gächter, S. & Fehr, E. (1999) Collective action as a social exchange. *Journal of Economic Behavior and Organization* 39(4):341–69. [aHG]
Gadagkar, R. (1991) On testing the role of genetic asymmetries created by haplodiploidy in the evolution of eusociality in the Hymenoptera. *Journal of Genetics* 70(1):1–31. [aHG]
Gauthier, D. P. (1986) *Morals by agreement.* Clarendon Press. [PD]
Geary, D. C. (2005) *The origin of mind: Evolution of brain, cognition, and general intelligence.* American Psychological Association. [rHG]
Ghiselin, M. T. (1974) *The economy of nature and the evolution of sex.* University of California Press. [aHG]
Gigerenzer, G. (2002) The adaptive toolbox. In: *Bounded rationality: The adaptive toolbox*, ed. G. Gigerenzer & R. Selten. MIT Press. [ODJ]
Gigerenzer, G. & Selten, R. (2001) *Bounded rationality.* MIT Press. [aHG]
Gilbert, G. N. & Troitzsch, K.G. (1999) *Simulation for the social scientist.* Open University Press. [RAB]
Gilovich, T., Vallone, R. & Tversky, A. (1985) The hot hand in basketball: On the misperception of random sequences. *Journal of Personality and Social Psychology* 17:295–314. [aHG]
Gintis, H. (1972) A radical analysis of welfare economics and individual development. *Quarterly Journal of Economics* 86(4):572–99. [aHG]
(1975) Welfare economics and individual development: A reply to Talcott Parsons. *Quarterly Journal of Economics* 89(2):291–302. [aHG]
(1980) Theory, practice, and the tools of communicative discourse. *Socialist Review* 50:189–232. [rHG]
(2000a) A great book with a fatally flawed model of human behavior. Review of R. D. Alexander's *The biology of moral systems.* Retrieved July 15, 2006, from http://www.amazon.com/gp/cdp/member-reviews/A2U0XHQB7MMH0E/102-4258424-8778549?ie=UTF8&display=public&page=7. [TG]
(2000b) A masterful historical and interpretive success. Review of Ullica Segerstråle's *Defenders of the truth.* Retrieved July 15, 2006, from: http://www.amazon.com/gp/cdp/member-reviews/A2U0XHQB7MMH0E/102-4258424-8778549?ie=UTF8&display=public&page=8. [TG]
(2000c) *Game theory evolving.* Princeton University Press. [aHG]
(2000d) Strong reciprocity and human sociality. *Journal of Theoretical Biology* 206:169–79. [RMB, arHG, RN]
(2003a) Solving the puzzle of human prosociality. *Rationality and Society* 15(2):155–87. [arHG]
(2003b) The hitchhiker's guide to altruism: Genes, culture, and the internalization of norms. *Journal of Theoretical Biology* 220(4):407–18. [arHG]
(2005a) A bioeconomic masterpiece. Review of Paul Seabright's *The company of strangers.* Retrieved July 15, 2006, from: http://www.amazon.com/gp/cdp/member-reviews/A2U0XHQB7MMH0E/102-4258424-8778549?ie=UTF8&display=public&page=3. [TG]
(2005b) Behavioral game theory and contemporary economic theory. *Analyze & Kritik* 27(1):48–72. [aHG]
(2005c) Outdated in detail, still a telling critique in broad outline. Review of Philip Kitcher's *Vaulting ambition.* Retrieved July 15, 2006, from http://www.amazon.com/gp/cdp/member-reviews/A2U0XHQB7MMH0E/102-4258424-8778549?ie=UTF8&display=public&page=1. [TG]
(2006a) Behavioral ethics meets natural justice. *Politics, Philosophy and Economics* 5(1):5–32. [PD]
(2006b) Moral skepticism defended. Review of Richard Joyce's *The evolution of morality.* Retrieved July 15, 2006, from: http://www.amazon.com/gp/cdp/member-reviews/A2U0XHQB7MMH0E/102-4258424-8778549?ie=UTF8&display=public&page=1. [TG]
(2006c) The emergence of a price system from decentralized bilateral exchange. *Contributions to Theoretical Economics* 6(1): Article 13. [aHG]
(in press a) The dynamics of general equilibrium. *The Economic Journal.* [aHG]
(in press b) The evolution of private property. *Journal of Economic Behavior and Organization.* [arHG]
Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003) Explaining altruistic behavior in humans. *Evolution and Human Behavior* 24:153–72. [RMB]
(2005b) Moral sentiments and material interests: Origins, evidence, and consequences. In: *Moral sentiments and material interests: The foundations of cooperation in economic life*, ed. H. Gintis, S. Bowles, R. Boyd & E. Fehr. MIT Press. [LA]
Gintis, H., Bowles, S., Boyd, R. & Fehr, E., ed. (2005a) *Moral sentiments and material interests: On the foundations of cooperation in economic life.* MIT Press. [LA, arHG]
Gintis, H., Smith, E. A. & Bowles, S. (2001) Costly signaling and cooperation. *Journal of Theoretical Biology* 213:103–19. [rHG]
Glassman, R. B. (1973) Persistence and loose coupling in living systems. *Behavioral Science* 18:83–98. [RBG]
(2000) A "theory of relativity" for cognitive elasticity of time and modality dimensions supporting constant working memory capacity: Involvement of harmonics among ultradian clocks? *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 24:163–82. [RBG]
(2004) Good behavioral science has room for theology: Any room for God? *Behavioral and Brain Sciences* 27:737–38. [RBG]
(2005) The epic of personal development and the mystery of small working memory. *Zygon/Journal of Religion and Science* 40:107–30. [RBG]
(2006) Metaphysics of money: A special case of emerging autonomy in evolving subsystems. *Behavioral and Brain Sciences* 29(2):186–87. [RBG]
Glassman, R. B., Packel, E. W. & Brown, D. L. (1986) Green beards and kindred spirits: A preliminary mathematical model of altruism toward nonkin who bear similarities to the giver. *Ethology and Sociobiology* 7:107–15. [RBG]
Glassman, R. B. & Smith, A. (1988) Neural spare capacity and the concept of diaschisis: Functional and evolutionary models. In: *Brain injury and recovery: Theoretical and controversial issues*, ed. S. Finger, T. E. LeVere, C. R. Almli & D. G. Stein, pp. 45–69. Plenum Press. [RBG]
Glassman, R. B. & Wimsatt, W. C. (1984) Evolutionary advantages and limitations of early plasticity. In: *Early brain damage, vol. 1: Research orientations and clinical observations*, ed. C. R. Almli & S. Finger, pp. 35–58. Academic Press. [RBG]
Glimcher, P. W. (2003) *Decisions, uncertainty, and the brain: The science of neuroeconomics.* MIT Press. [aHG]
Glimcher, P. W., Dorris, M. C. & Bayer, H. M. (2005) *Physiological utility theory and the neuroeconomics of choice.* Center for Neural Science, New York University. [aHG]
Gneezy, U. (2005) Deception: The role of consequences. *American Economic Review* 95(1):384–94. [aHG]
Gold, N. & Sugden, R. (in press) Theories of team agency. In: *Rationality and commitment*, ed. F. Peter & H. B. Schmid. Oxford University Press. [AMC]
Goldstein, E. B. (2005) *Cognitive psychology: Connecting mind, research, and everyday experience.* Wadsworth. [aHG]
Gow, D. W. (2003) How representations help define computational problems. *Journal of Phonetics* 31:487–93. [DWG]
Grafen, A. (1999) Formal Darwinism, the individual-as-maximizing-agent analogy, and bet-hedging. *Proceedings of the Royal Society B* 266:799–803. [aHG]
(2000) Developments of Price's equation and natural selection under uncertainty. *Proceedings of the Royal Society B* 267:1223–27. [aHG]
(2002) A first formal link between the Price equation and an optimization program. *Journal of Theoretical Biology* 217:75–91. [aHG]
(2006) Optimization of inclusive fitness. *Journal of Theoretical Biology* 238:541–63. [rHG]
Granovetter, M. (2003) Ignorance, knowledge, and outcomes in a small world. *Science* 301:773–74. [RAB]
Grether, D. & Plott, C. (1979) Economic theory of choice and the preference reversal phenomenon. *American Economic Review* 69(4):623–38. [aHG]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Grice, H. P. (1975) Logic and conversation. In: *The logic of grammar*, ed. D. Davidson & G. Harman, pp. 64–75. Dickenson. [aHG]
Gruber, J. & Koszegi, B. (2001) Is addiction rational? Theory and evidence. *Quarterly Journal of Economics* 116(4):1261–1305. [aHG]
Grusec, J. E. & Kuczynski, L. (1997) *Parenting and children's internalization of values: A handbook of contemporary theory.* Wiley. [aHG]
Gunnthorsdottir, A., McCabe, K. & Smith, V. (2002) Using the Machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology* 23:49–66. [aHG]
Hagen, E. H. & Hammerstein, P. (2006) Game theory and human evolution: A critique of some recent interpretations of experimental games. *Theoretical Population Biology* 69:339–48. [RMB, TG aHG]
Hahn, M. W. & Bentley, R. A. (2003) Drift as a mechanism for cultural change: An example from baby names. *Proceedings of the Royal Society B* 270:S120–203. [RAB]
Haig, D. (2000) Genomic imprinting, sex-biased dispersal, and social behavior. *Annals of the New York Academy of Sciences* 907:149–63. [MEP]
(2003) On intrapersonal reciprocity. *Evolution and Human Behavior* 24:418–25. [MEP]
Haldane, J. B. S. (1932) *The causes of evolution.* Longmans, Green. [aHG]
Hamilton, W. D. (1963) The evolution of altruistic behavior. *American Naturalist* 96:354–56. [aHG]
(1964) The genetical evolution of social behavior, I & II. *Journal of Theoretical Biology* 7:1–16, 17–52. [rHG, MEP]
Hammerstein, P. (1996) Darwinian adaptation, population genetics and the streetcar theory of evolution. *Journal of Mathematical Biology* 34:511–32. [aHG]
(2003) Why is reciprocity so rare in social animals? A protestant appeal. In: *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, pp. 83–93. The MIT Press. [aHG]
Hammerstein, P. & Selten, R. (1994) Game theory and evolutionary biology. In: *Handbook of game theory with economic applications*, ed. R. J. Aumann & S. Hart, pp. 929–993. Elsevier. [aHG]
Hammond, K. R. & Stewart, T. R. (2001) *The essential Brunswik: Beginnings, explications, applications.* Oxford University Press. [KRH]
Handwerker, W. P. (1989) The origin and evolution of culture. *American Anthropologist* 91:313–26. [DPK]
(2001) *Quick ethnography: A guide to rapid multi-method research.* AltaMira Press. [DPK]
Harker, R., Mahar, C. & Wilkes, C. (1990) *An introduction to the work of Pierre Bourdieu.* Macmillan. [SC]
Harsanyi, J. C. (1967) Games with incomplete information played by Bayesian players, Parts I, II, and III. *Behavioral Science* 14:159–82, 320–34, 486–502. [aHG]
Harsanyi, J. C. & Selten, R. (1988) *A general theory of equilibrium selection in games.* MIT Press. [AMC]
Harte, J. M., Westenberg, M. R. & van Someren, M. (1994) Process models of decision making. *Acta Psychologica* 87:95–120. [MS-M]
Haucap, J. & Just, T. (2003) Not guilty? Another look at the nature and nurture of economics students. Discussion Paper 8, University of the Federal Armed Forces, Hamburg, Germany. Available at: www.ruhr-uni-bochum.de/wettbewerb/dlls/forschung/paper8.pdf [GA]
Hayes, B. (2006) The semicolon wars. *American Scientist* 94:299–303. [ABM]
Hechter, M. & Kanazawa, S. (1997) Sociological rational choice. *Annual Review of Sociology* 23:199–214. [aHG]
Heiner, R. A. (1983) The origin of predictable behavior. *American Economic Review* 73(4):560–95. [aHG]
Henrich, J. (1997) Market incorporation, agricultural change and sustainability among the Machiguenga Indians of the Peruvian Amazon. *Human Ecology* 25:319–51. [aHG]
(2001) Cultural transmission and the diffusion of innovations. *American Anthropologist* 103:992–1013. [aHG]
Henrich, J. & Boyd, R. (1998) The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior* 19:215–42. [aHG]
(2001) Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology* 208:79–89. [RAB]
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. & McElreath, R. (2004) Overview and synthesis. In: *Foundations of human sociality. Economic experiments and ethnographic evidence from fifteen small-scale societies*, ed. J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr & H. Gintis, pp. 8–54. Oxford University Press. [RN]
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E. & Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Smith Henrich, N., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q. & Tracer, D. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*. 28(6):795–815. [aHG, JWP]
Henrich, J. & Gil-White, F. (2001) The evolution of prestige: Freely conferred status as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* 22:165–96. [aHG, RAB]
Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J.C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. & Ziker, J. (2006) Costly punishment across human societies. *Science* 312:1767–70. [RAB]
Henrich, J. P. (2004) *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies.* Oxford University Press. [PD]
Herrnstein, R. J. (1961) Relative and absolute strengths of responses as a function of frequency of reinforcement. *Journal of Experimental Analysis of Animal Behavior* 4:267–72. [aHG]
Herrnstein, R., Laibson, D. & Rachlin, H. (1997) *The matching law: Papers on psychology and economics.* Harvard University Press. [aHG]
Herzog, H. A., Bentley, R. A. & Hahn, M. W. (2004) Random drift and large shifts in popularity of dog breeds. *Proceedings of the Royal Society B* 271:S353–56. [RAB]
Hewlett, B., De Silvestri, A. & Guglielmino, C. R. (2002) Semes and genes in Africa. *Current Anthropology* 43:313–21. [AM]
Hilton, D. J. (1995) The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin* 118(2):248–71. [aHG]
(2003) Psychology and the financial markets: Applications to understanding and remedying irrational decision-making. In: *The psychology of economic decisions. Vol. 1: Rationality and well-being*, ed. I. Brocas & J. D. Carrillo, pp. 273–97. Oxford University Press. [KES]
Hirsch, P., Michaels, S. & Friedman, R. (1990) Clean models vs. dirty hands: Why economics is different from sociology. In: *Structures of capital: The social organization of the economy*, ed. S. Zukin & P. DiMaggio, pp. 39–56. Cambridge University Press. [aHG]
Hirschfeld, L. A. (1996) *Race in the making.* MIT Press. [ABM]
Hochberg, L. R., Serruya, M. D., Friehs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D. & Donoghue, J. P. (2006) Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature* 442:164–71. [JIK]
Hodgson, G. M. (2001) *How economics forgot history: The problem of historical specificity in social science.* Routledge. [GMH]
(2004) *The evolution of institutional economics: Agency, structure and Darwinism in American institutionalism.* Routledge. [GMH]
Hoffman, M. B. & Goldsmith, T. H. (2004) The biological roots of punishment. *Ohio State Journal of Criminal Law* 1:627–41. Available at: http://heinonline.org/HOL/Page?handle=hein.journals/osjcl1&id=635&collection=cjjournals. [ODJ]
Hoffmaster, B. (1993) Can ethnography save the life of medical ethics? *Social Science and Medicine* 35(12):1421–31. [PD]
Holden, C. J. (2002) Bantu language trees reflect the spread of farming across Sub-Saharan Africa: A maximum-parsimony analysis. *Proceedings of the Royal Society of London Series B* 269:793–99. [aHG, AM]
Holden, C. J. & Mace, R. (2003) Spread of cattle led to the loss of matrilineal descent in Africa: A coevolutionary analysis. *Proceedings of the Royal Society of London Series B* 270:2425–33. [aHG]
Holland, J. H. (1975) *Adaptation in natural and artificial systems.* University of Michigan Press. [rHG]
(1995) *Hidden order: How adaptation builds complexity.* Helix Books. [GMH]
Hollis, M. (1983) Rational preferences. *The Philosophical Forum* 14:246–62. [GA]
(1998) *Trust within reason.* Cambridge University Press. [AMC]
Horvitz, J. C. (2000) Mesolimbocortical and nigrostriatal dopamine responses to salient non-reward events. *Neuroscience* 96:651–56. [DJZ]
Huang, C.-F. & Litzenberger, R. H. (1988) *Foundations for financial economics.* Elsevier. [aHG]
Hurst, L. D., Atlan, A. & Bengtsson, B. O. (1996) Genetic conflicts. *Quarterly Review of Biology* 71(3):317–64. [JWP]
Huxley, J. S. (1955) Evolution, cultural and biological. *Yearbook of Anthropology*, pp. 2–25. [aHG]
Huxley, T. H. & Martin, H. N. (1888) *A course of elementary instruction in practical biology*, 2nd ed. Macmillan (original work published in 1875). [JF]
Irons, W. G. (1979) Natural selection, adaptation, and human social behavior. In: *Evolutionary biology and human behavior*, ed. N. Chagnon & W. Irons, pp. 4–39. Duxbury Press. [RLB]
Jablonka, E. & Lamb, M. J. (1995) *Epigenetic inheritance and evolution: The Lamarckian case.* Oxford University Press. [aHG]
Jackendoff, R. (1996) How language helps us think. *Pragmatics and Cognition* 4:1–34. [KES]
Jacobs, R. C. & Campbell, D. T. (1961) The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *Journal of Abnormal and Social Psychology* 62:649–58. [AM]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

James, C. D., Hoffman, M. T., Lightfoot, D. C., Forbes, G. S. & Whitford, W. G. (1994) Fruit abortion in *Yucca elata* and its implications for the mutualistic association with yucca moths. *Oikos* 69:207–16. [RN]
James, W. (1880) Great men, great thoughts, and the environment. *Atlantic Monthly* 46:441–59. [aHG]
Janssen, M. (2001) Rationalising focal points. *Theory and Decision* 50:119–48. [AMC]
Jaynes, E. T. (2003) *Probability theory: The logic of science.* Cambridge University Press. [aHG]
Jeffrey, R. (1974) Preferences among preferences. *Journal of Philosophy* 71:377–91. [KES]
Johnson, E. J., Schulte-Mecklenbeck, M. & Willemsen, M. (under review) Process models deserve process data. *Psychological Review*. [MS-M]
Johnson, P. (1987) *A history of the Jews*. Harper & Row. [RBG]
Jones, O. D. (1997) Evolutionary analysis in law: An introduction and application to child abuse. *North Carolina Law Review* 75:1117–242. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=611961. [ODJ]
(1999) Sex, culture, and the biology of rape: Toward explanation and prevention. *California Law Review* 87:827–942. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=611908. [ODJ]
(2001) Time-shifted rationality and the law of law's leverage: Behavioral economics meets behavioral biology. *Northwestern Law Review* 95:1141–206. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=249419. [ODJ]
(2004) Law, evolution and the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359:1697–707. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=692742. [ODJ]
Jones, O. D. & Goldsmith, T. H. (2005) Law and behavioral biology. *Columbia Law Review* 105:405–502. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=688619. [ODJ]
Kagan, S. (1991) *The limits of morality.* Oxford University Press. [PD]
Kahneman, D. & Frederick, S. (2002) Representativeness revisited: Attribute substitution in intuitive judgment. In: *Heuristics and biases: The psychology of intuitive judgment*, ed. T. Gilovich, D. Griffin & D. Kahneman, pp. 49–81. Cambridge University Press. [KES]
(2005) A model of heuristic judgment. In: *The Cambridge handbook of thinking and reasoning*, ed. K. J. Holyoak & R. G. Morrison, pp. 267–93. Cambridge University Press. [KES]
Kahneman, D., Slovic, P. & Tversky, A., eds. (1982) *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press. [JT]
Kahneman, D. & Tversky, A. (1979) Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–91. [arHG, JT]
(2000) *Choices, values, and frames.* Cambridge University Press. [aHG]
Kahneman, D., Slovic, P. & Tversky, A., eds. (1982) *Judgment under uncertainty: Heuristics and biases.* Cambridge University Press. [aHG]
Keller, L. M., ed. (1999) *Levels of selection in evolution*. Princeton University Press. [JWP]
Kerr, B. & Godfrey-Smith, P. (2002) Individualist and multi-level perspectives on selection in structured populations. *Biology and Philosophy* 17:477–517. [rHG]
Kiers, E. T., Rousseau, R. A., West, S. A. & Denison, R. F. (2003) Host sanctions and the legume-rhizobium mutualism. *Nature* 425:78–81. [RN]
Kihlstrom, J. F. (2006) Does neuroscience constrain social-psychological theory? *Dialogue* [Society for Personality & Social Psychology] 21(1):16–17, 32. [JIK]
Kitcher, P. (1989) Explanatory unification and the causal structure of the world. In: *Scientific explanation*, ed. P. Kitcher & W. Salmon. University of Minnesota Press. [SC]
(1999) Unification as a regulative ideal. *Perspectives on Science* 7:337–48. [SC]
Kiyonari, T., Tanida, S. & Yamagishi, T. (2000) Social exchange and reciprocity: Confusion or a heuristic? *Evolution and Human Behavior* 21:411–27. [aHG]
Kohler, T. A. & Gumerman, G. J., eds. (2000) *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes.* Oxford University Press. [AM]
Kollock, P. (1997) Transforming social dilemmas: Group identity and cooperation. In: *Modeling Rational and Moral Agents*, ed. P. Danielson. Oxford University Press. [aHG]
Konner, M. (2003) *Unsettled: An anthropology of the Jews*. Viking Penguin. [RBG]
Krantz, D. H. (1991) From indices to mappings: The representational approach to measurement. In: *Frontiers of Mathematical Psychology*, ed. D. Brown & J. Smith, pp. 1-52. Cambridge University Press. [aHG]
Krantz, D. L. (2001) Reconsidering history of psychology's borders. *History of Psychology* 4:182–94. [RBG]
Krebs, J. R. & Davies, N. B. (1997a) *Behavioural ecology: An evolutionary approach*, 4th edition. Blackwell Science. [aHG]
(1997b) The evolution of behavioural ecology. In: *Behavioural ecology: An evolutionary approach*, 4th edition, ed. J. R. Krebs & N. B. Davies, pp. 3–12. Blackwell Scientific. [TG]
Kreps, D. M. (1990) *A course in microeconomic theory.* Princeton University Press. [aHG]
Krueger, J. I. (1998) The bet on bias: A foregone conclusion? *Psycoloquy* 9(46). http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?9.46. [JIK]
(2003) Wanted: A reconciliation of rationality with determinism. *Behavioral and Brain Sciences* 26:168–69. [JIK]
(2004) Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences* 27(3):313–27. [aHG]
Kuhn, T. (1962) *The structure of scientific revolutions*. University of Chicago Press. [JIK, aHG]
(1970) *The structure of scientific revolutions*, 2nd edition. University of Chicago Press. [SC]
Kummel, M. & Salant, S. W. (2006) The economics of mutualisms: Optimal utilization of mycorrhizal mutualistic partners by plants. *Ecology* 87(4):892–902. [RN]
Kurz, M. (1997) Endogenous economic fluctuations and rational beliefs: A general perspective. In: *Endogenous economic fluctuations: Studies in the theory of rational beliefs*, ed. M. Kurz, pp. 1–37. Springer-Verlag. [aHG]
Kurzban, R. & Houser, D. (2005) Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations. *Proceedings of the National Academy of Sciences USA* 102(5):1083–807. [PD]
Laibson, D. (1997) Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 112(2):443–77. [arHG]
Laibson, D., Choi, J. & Madrian, B. (2004) Plan design and 401(k) savings outcomes. *National Tax Journal* 57:275–98. [arHG]
Laland, K. N. (2004) Social learning strategies. *Learning and Behavior* 32:4–14. [AM]
Laland, K. N. & Brown, G. R. (2002) *Sense and nonsense: Evolutionary perspectives on human behaviour.* Oxford University Press. [AM]
Laland, K. N., Kumm, J. & Feldman, M. W. (1995) Gene-culture coevolutionary theory – A test-case. *Current Anthropology* 36:131–56. [AM]
Lansing, J. S. (2006) *Perfect order: Recognizing complexity in Bali*. Princeton University Press. [RAB]
Lashley, K. S. (1930) Basic neural mechanisms in behavior. *Psychological Review* 37:1–24. [RBG]
Laszlo, E. (1973) *Introduction to systems philosophy.* Harper Torchbooks [RBG]
Lea, S. E. G. & Webley, P. (2006) Money as tool, money as drug: The biological psychology of a strong incentive. *Behavioral and Brain Sciences* 29(2):161–209. [GA]
Level Playing Field Institute. (2006) *HOW-FAIR executive summary.* http://www.lpfi.org/workplace/howfair.shtml. [RN]
Levy, N. (2002) *Sartre*. Oneworld. [SC]
Lewontin, R. C. (1974) *The genetic basis of evolutionary change.* Columbia University Press. [aHG]
(1990) The evolution of cognition. In: *An invitation to cognitive science, vol. 3*, ed. D. H. Osherson. MIT Press. [DJZ]
Liberman, U. (1988) External stability and ESS criteria for initial increase of a new mutant allele. *Journal of Mathematical Biology* 26:477–85. [aHG]
Lichtenstein, S. & Slovic, P. (1971) Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology* 89:46–55. [aHG]
Lieberman, D., Tooby, J. & Cosmides, L. (2007) The architecture of human kin detection. *Nature* 445(7129):727–31. [JT]
Lipo, C. P., Madsen, M. E., Dunnell, R. C. & Hunt, T. (1997) Population structure, cultural transmission, and frequency seriation. *Journal of Anthropological Archaeology* 16:301–33. [RAB]
Lipo, C. P., O'Brien, M. J., Collard, M. & Shennan, S., eds. (2006) *Mapping our ancestors: Phylogenetic approaches in anthropology and prehistory.* Aldine. [AM]
Loomes, G. & Sugden, R. (1982) Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal* 92:805–24. [aHG]
Luce, R. D. (2000) *Utility of gains and losses: Measurement-theoretical and experimental approaches*. Erlbaum. [RLB]
Lumsden, C. J. & Wilson, E. O. (1981) *Genes, mind, and culture: The coevolutionary process.* Harvard University Press. [aHG]
(1983) *Promethean fire: Reflections on the origin of mind*. Harvard University Press. [RBG]
Mace, R. & Holden, C. J. (2005) A phylogenetic approach to cultural evolution. *Trends in Ecology and Evolution* 20:116–21. [AM]
Mace, R. & Pagel, M. (1994) The comparative method in anthropology. *Current Anthropology* 35:549–64. [aHG]
Mandeville, B. (1705/1924) *The fable of the bees: Private vices, publick benefits.* Clarendon Press. [aHG]
Margulis, L. (1970) *Origin of eukaryotic cells.* Yale University Press. [JWP]
Markman, A. B. (1999) *Knowledge representation*. Erlbaum. [ABM]
Marr, D. (1982) *Vision.* W. H. Freeman. [DWG]
Mas-Colell, A., Whinston, M. D. & Green, J. R. (1995) *Microeconomic theory.* Oxford University Press. [rHG]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Maynard Smith, J. (1976) Group selection. *Quarterly Review of Biology* 51:277–83. [aHG]
(1982) *Evolution and the theory of games.* Cambridge University Press. [aHG]
Maynard Smith, J. & Szathmáry, E. (1995/1997) *The major transitions in evolution*. Freeman/Oxford University Press. [arHG, JWP]
Mayr, E. (1988) *Toward a new philosophy of biology: Observations of an evolutionist.* Harvard University Press. [GMH]
(1996) The autonomy of biology: The position of biology among the sciences. *Quarterly Review of Biology* 71:97–106. [LA]
(1997) The objects of selection. *Proceedings of the National Academy of Sciences* 94:2091–94. [rHG]
(2004) *What makes biology unique? Considerations on the autonomy of a scientific discipline*. Cambridge University Press. [LA]
Mazur, A. (2005) *Biosociology of dominance and deference*. Rowman and Littlefield. [GEW]
Mazur, J. E. (2002) *Learning and behavior*. Prentice-Hall. [aHG]
McCain, R. A. (2003) Specifying agents: Probabilistic equilibrium with non-self-interested motives. Paper presented at the Ninth International Conference on Computing in Economics and Finance, University of Washington, Seattle, July 11–13, 2003. Available at: http://william-king.www.drexel.edu/top/eco/agents.pdf. [RAM]
McClure, S. M., Laibson, D. I., Loewenstein, G. & Cohen, J. D. (2004) Separate neural systems value immediate and delayed monetary rewards. *Science* 306(5695):503–507. [arHG]
McKelvey, R. D. & Palfrey, T. R. (1992) An experimental study of the centipede game. *Econometrica* 60:803–36. [aHG]
(1995) Quantal response equilibria for normal form games. *Games and Economic Behavior* 10:6–38. [RAM]
McNamara, J. M., Houston, A. I. & Collins, E. J. (2001) Optimality models in behavioral biology. *Siam Review* 43:413–66. [TG]
Mead, M. (1963) *Sex and temperament in three primitive societies.* Morrow. [aHG]
Medin, D. L. & Atran, S. (2004) The native mind; Biological categorization and reasoning in development and across cultures. *Psychological Review* 111(4):960–83. [ABM]
Mehta, J., Starmer, C. & Sugden, R. (1994) The nature of salience: An experimental investigation of pure coordination games. *American Economic Review* 84:658–73. [AMC]
Meltzhoff, A. N. & Decety, J. (2003) What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London B* 358:491–500. [aHG]
Menzies, G. D. & Zizzo, D. J. (2006) Rational expectations. Social Science Research Network Discussion Paper. http://ssrn.com/abstract=913374. [DJZ]
Mesoudi, A. & O'Brien, M. J. (submitted) The cultural transmission of Great Basin projectile point technology: Experimental and computer simulations. [AM]
Mesoudi, A., Whiten, A. & Laland, K. N. (2004) Is human cultural evolution Darwinian? Evidence reviewed from the perspective of *The Origin of Species*. *Evolution* 58:1–11. [AM]
(2006) Towards a unified science of cultural evolution. *Behavioral and Brain Sciences* 29(4):329–83. [aHG, AM]
Michod, R. E. (1999) *Darwinian dynamics: Evolutionary transitions in fitness and individuality*. Princeton University Press. [JWP]
Miller, B. L., Darby, A., Benson, D. F., Cummings, J. L. & Miller, M. H. (1997) Aggressive, socially disruptive and antisocial behaviour associated with fronto-temporal dementia. *British Journal of Psychiatry* 170:150–54. [aHG]
Miller, G. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2):81–97. [DPK]
Miller, W. R. (2003) Comments on Ainslie and Monterosso. In: *Choice, behavioural economics, and addiction*, ed. R. Vuchinich & N. Heather, pp. 62–66. Pergamon Press. [GA]
Milner, P. M. (1970) *Physiological psychology*. Holt, Rinehart & Winston. [RBG]
Mirowski, P. (1989) *More heat than light: Economics as social physics, physics as nature's economics*. Cambridge University Press. [LA]
Molenaar, P. C. M. (2004) A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. Focus article. *Measurement* 2:201–18. [RLB]
Moll, J., Zahn, R., di Oliveira-Souza, R., Krueger, F. & Grafman, J. (2005) The neural basis of human moral cognition. *Nature Neuroscience* 6:799–809. [aHG]
Montague, P. R. & Berns, G. S. (2002) Neural economics and the biological substrates of valuation. *Neuron* 36:265–84. [aHG]
Moore, Jr., B. (1978) *Injustice: The social bases of obedience and revolt.* Sharpe. [LA, aHG]
Moran, P. A. P. (1964) On the nonexistence of adaptive topographies. *Annals of Human Genetics* 27:338–43. [aHG]
Morowitz, H. (2002) *The emergence of everything: How the world became complex*. Oxford University Press. [rHG, LA]
Nagel, E. (1961) *The structure of science.* Routledge and Hackett. [GMH]
Nakamaru, M. & Iwasa, Y. (2006) The coevolution of altruism and punishment: Role of the selfish punisher. *Journal of Theoretical Biology* 240:475–88. [TG]
Neiman, F. D. (1995) Stylistic variation in evolutionary perspective – Inferences from decorative diversity and interassemblage distance in Illinois woodland ceramic assemblages. *American Antiquity* 60:7–36. [RAB, AM]
Nesse, R. M. & Williams, G. C. (1994/1996) *Why we get sick: The new science of Darwinian medicine*. Random House/Vintage Books. [ODJ, GEW]
Neuberg, S. L., Cialdini, R. B., Brown, S. L., Luce, C., Sagarin, B. & Lewis, B. P. (1997) Does empathy lead to anything more than superficial helping? Comment on Batson et al. (1997). *Journal of Personality and Social Psychology* 73:510–16. [RMB]
Newman, M. E. J., Barabasi, A.-L. & Watts, D. J., eds. (2006) *The structure and dynamics of networks.* Princeton University Press. [aHG]
Nisbett, R. E. & Cohen, D. (1996) *Culture of honor: The psychology of violence in the South.* Westview Press. [aHG]
Noë, R. (2001) Biological markets: Partner choice as the driving force behind the evolution of cooperation. In: *Economics in nature. Social dilemmas, mate choice and biological markets*, ed. R. Noë, J. A. R. A. M. van Hooff & P. Hammerstein, pp. 93–118. Cambridge University Press. [RN]
Nowak, M. A. & Sigmund, K. (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* 364:56–58. [AMC]
Nozick, R. (1993) *The nature of rationality.* Princeton University Press. [KES]
O'Brien, M. J. & Lyman, R. L. (2000) *Applying evolutionary archaeology.* Kluwer Academic. [aHG]
(2003) *Cladistics and archaeology.* University of Utah Press. [AM]
Odling-Smee, F. J., Laland, K. N. & Feldman, M. W. (2003) *Niche construction: The neglected process in evolution.* Princeton University Press. [arHG]
O'Donoghue, T. & Rabin, M. (2001) Choice and procrastination. *Quarterly Journal of Economics* 116(1):121–60. [aHG]
O'Hara, E. (2004) How neuroscience might advance the law. *Philosophical Transactions of the Royal Society B: Biological Sciences* 359:1677–84. Available at: http://www.journals.royalsoc.ac.uk/(0i3l2rylficf2cbdgnwn442y)/app/home/contribution.asp?referrer=parent&backto=issue,4,16;journal,23,125;linkingpublicationresults,1:102022,1. [ODJ]
Ohmagari, K. & Berkes, F. (1997) Transmission of indigenous knowledge and bush skills among the Western James Bay Cree women of subarctic Canada. *Human Ecology* 25:197–222. [AM]
Olson, M. (1965) *The logic of collective action: Public goods and the theory of groups.* Harvard University Press. [aHG]
Omark, D. R., Strayer, F. F. & Freedman, D. G. (1980) *Dominance relations: An ethological view of human conflict and social interaction.* Garland Press. [GEW]
Ormerod, P. (1998) *Butterfly economics*. Faber and Faber. [RAB]
(2005) *Why most things fail*. Faber and Faber. [RAB]
Ostrom, E., Walker, J. & Gardner, R. (1992) Covenants with and without a sword: Self-governance is possible. *American Political Science Review* 86(2):404–17. [aHG]
Owens, P. F. (2006) Where is behavioural ecology going? *Trends in Ecology and Evolution* 21:356–61. [TG]
Panksepp, J. (1998) *Affective neuroscience.* Oxford University Press. [GEW]
Parker, A. J. & Newsome, W. T. (1998) Sense and the single neuron: Probing the physiology of perception. *Annual Review of Neuroscience* 21:227–77. [aHG]
Parsons, T. (1964) Evolutionary universals in society. *American Sociological Review* 29(3):339–57. [aHG]
(1967) *Sociological theory and modern society.* Free Press. [aHG]
Parsons, T. & Shils, S. (1951) *Toward a general theory of action.* Harvard University Press. [aHG, JT]
Payne, J. W., Bettman, J. R. & Johnson, E. J. (1993) *The adaptive decision maker*. Cambridge University Press. [MS-M]
Pearce, D. (1984) Rationalizable strategic behavior and the problem of perfection. *Econometrica* 52:1029–50. [aHG]
Pellmyr, O. & Huth, C. J. (1994) Evolutionary stability of mutualism between yuccas and yucca moths. *Nature* 372:257–60. [RN]
Perner, J. (1991) *Understanding the representational mind.* MIT Press. [KES]
Pinker, S. (2002) *The blank slate: The modern denial of human nature.* Viking. [aHG]
Plotkin, H. C. (1994) *Darwin machines and the nature of knowledge: Concerning adaptations, instinct and the evolution of intelligence.* Penguin. [GMH]
Plott, C. R. (1979) The application of laboratory experimental methods to public choice. In: *Collective decision making: Applications from public choice theory*, ed. C. S. Russell, pp. 137–60. Johns Hopkins University Press. [aHG]
Polya, G. (1990) *Patterns of plausible reasoning.* Princeton University Press. [aHG]
Pool, I. de S. & Kochen, M. (1978) Contacts and influence. *Social Networks* 1:5–51. [RAB]
Popper, K. (1959) *The logic of scientific discovery*. Basic Books. [JWP]
(1979) *Objective knowledge: An evolutionary approach.* Clarendon. [aHG]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Poundstone, W. (1992) *Prisoner's Dilemma.* Doubleday. [aHG]
Povinelli, D. J. & Bering, J. M. (2002) The mentality of apes revisited. *Current Directions in Psychological Science* 11(4):115–19. [KES]
Povinelli, D. J. & Giambrone, S. (2001) Reasoning about beliefs: A human specialization? *Child Development* 72:691–95. [KES]
Power, T. G. & Chapieski, M. L. (1986) Childrearing and impulse control in toddlers: A naturalistic investigation. *Developmental Psychology* 22:271–75. [aHG]
Premack, D. (1959) Toward empirical behavior laws, I. Positive reinforcement. *Psychological Review* 66:219–34. [GA]
Price, G. R. (1970) Selection and covariance. *Nature* 227:520–21. [rHG]
Rabin, M. (1993) Incorporating fairness into game theory and economics. *American Economic Review* 83:1281–1302. [AMC]
(2002) Inference by believers in the law of small numbers. *Quarterly Journal of Economics* 117(3):775–816. [aHG]
Radin, M. (1996) *Contested commodities*. Harvard University Press. [SC]
Real, L. A. (1991) Animal choice behavior and the evolution of cognitive architecture. *Science* 253:980–86. [aHG]
Real, L. A. & Caraco, T. (1986) Risk and foraging in stochastic environments. *Annual Review of Ecology and Systematics* 17:371–90. [aHG]
Redgrave, P., Prescott, T. J. & Gurney, K. (1999a) Is the short-latency dopamine response too short to signal reward error? *Trends in Neurosciences* 22:146–51. [DJZ]
(1999b) The basal ganglia: A vertebrate solution to the selection problem? *Neuroscience* 89:1009–23. [DJZ]
Redlawsk, D. (2004) What voters do: Information search during election campaigns. *Political Psychology* 25:595–609. [MS-M]
Resch, R. P. (1992) *Althusser and the renewal of Marxist social theory*. University of California Press. [SC]
Richerson, P. J. & Boyd, R. (1998) The evolution of ultrasociality. In: *Indoctrinability, ideology and warfare*, ed. I. Eibl-Eibesfeldt & F. K. Salter, pp. 71–96. Berghahn Books. [aHG]
(2004) *Not by genes alone.* University of Chicago Press. [aHG]
Ridley, M. (2001) *The cooperative gene*. Simon & Schuster. [JWP]
Rivera, M. C. & Lake, J. A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature* 431:152–55. [aHG]
Rizzolatti, G., Fadiga, L., Fogassi, L. & Gallese, V. (2002) From mirror neurons to imitation: Facts and speculations. In: *The imitative mind: Development, evolution and brain bases*, ed. A. N. Meltzhoff & W. Prinz, pp. 247–66. Cambridge University Press. [aHG]
Robinson, P. H., Kurzban, R. & Jones, O. D. (in preparation) The origins of shared intuitions of justice. [ODJ]
Rode, C., Cosmides, L., Hell, W. & Tooby, J. (1999) When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition* 72:269–304. [JT]
Rogers, A. (1994) Evolution of time preference by natural selection. *American Economic Review* 84(3):460–81. [aHG]
Rohlfing, D. L.&. Oparin, A. I., eds. (1972) *Molecular evolution: Prebiological and biological*. Plenum Press. [JWP]
Rosen, S. (2005) *War and human nature*. Princeton University Press. [LA]
Rosenthal, R. W. (1981) Games of perfect information, predatory pricing and the chain-store paradox. *Journal of Economic Theory* 25:92–100. [aHG]
Rosenzweig, M. R., Breedlove, S. M. & Watson, N. V. (2005) *Biological psychology*, 4th edition. Sinauer. [RBG]
Rozin, P., Lowery, L., Imada, S. & Haidt, J. (1999) The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology* 76:574–86. [aHG]
Russo, J. E. (1978) Eye fixations can save the world: A critical evaluation and comparison with other information processing methodologies. In: *Advances in consumer research, vol. 5*, ed. H. K. Hunt, pp. 561–70. Association for Consumer Research. [MS-M]
Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. (2004) The evolution of cooperation. *Quarterly Review of Biology* 79(2):135–60. [RN]
Saffer, H. & Chaloupka, F. (1999) The demand for illicit drugs. *Economic Inquiry* 37(3):401–11. [aHG]
Salganik, M. J., Dodds, P. S. & Watts, D. J. (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311:854–56. [RAB]
Sally, D. (1995) Conversation and cooperation in social dilemmas. *Rationality and Society* 7(1):58–92. [aHG]
Samuels, R. & Stich, S. P. (2004) Rationality and psychology. In: *The Oxford handbook of rationality*, ed. A. R. Mele & P. Rawling, pp. 279–300. Oxford University Press. [KES]
Samuelson, L. (2001) Analogies, adaptation, and anomalies. *Journal of Economic Theory* 97:320–66. [rHG]
Sanchez, A. & Cuesta, J. A. (2005) Altruism may arise from individual selection. *Journal of Theoretical Biology* 235:233–40. [RMB]
Sanfey, A. G., Loewenstein, G., McClure, S. M. & Cohen, J. D. (2006) Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences* 10:108–16. [KES]
Savage, L. J. (1954) *The foundations of statistics.* Wiley. [aHG]
Schall, J. D. & Thompson, K. G. (1999) Neural selection and control of visually guided eye movements. *Annual Review of Neuroscience* 22:241–59. [aHG]
Schiffman, L. G. & Kanuk, L. L. (2004) *Consumer behavior*, 8th edition. Pearson/Prentice-Hall. [RBG]
Schkade, D. & Johnson, E. (1989) Cognitive processes in preference reversals. *Organizational Behavior and Human Decision Processes* 44:203–31. [MS-M]
Schrödinger, E. (1944) *What is life?: The physical aspect of the living cell.* Cambridge University Press. [aHG]
Schulkin, J. (2000) *Roots of social sensitivity and neural function.* MIT Press. [aHG]
Schultz, W., Dayan, P. & Montague, P. R. (1997) A neural substrate of prediction and reward. *Science* 275:1593–99. [aHG, DJZ]
Seeley, T. D. (1997) Honey bee colonies are group-level adaptive units. *The American Naturalist* 150:S22–S41. [aHG]
Segerstråle, U. (2001) *Defenders of the truth: The sociobiology debate.* Oxford University Press. [arHG]
Selten, R. (1993) In search of a better understanding of economic behavior. In: *The makers of modern economics, vol. 1*, ed. A. Heertje, pp. 115–39. Harvester Wheatsheaf. [aHG]
Shafir, E. and LeBoeuf, R. A. (2002) Rationality. *Annual Review of Psychology* 53:491–517. [aHG]
Shennan, S. (1997) *Quantifying archaeology.* Edinburgh University Press. [aHG]
Shennan S. J. & Wilkinson, J. R. (2001) Ceramic style change and neutral evolution: A case study from neolithic Europe. *American Antiquity* 66:577–94. [RAB, AM]
Sherif, M. (1936) *The psychology of social norms.* Harper. [AM]
Siegler, R., Deloache, J. & Eisenberg, N. (2003) *How children develop*. Worth. [RBG]
Simkin, M.V. & Roychowdhury, V.P. (2003) Read before you cite! *Complex Systems* 14:269. [RAB]
Simms, E. L., Taylor, D. L., Povich, J., Shefferson, R. P., Sachs, J. L., Urbina, M. & Tausczik, Y. (2006) An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proceedings of the Royal Society B-Biological Sciences* 273(1582):77–81. [RN]
Simon, H. (1972) Theories of bounded rationality. In: *Decision and organization*, ed. C. B. McGuire & R. Radner, pp. 161–76. Elsevier. [aHG]
(1982) *Models of bounded rationality.* MIT Press. [aHG]
Skibo, J. M. & Bentley, R. A. (2003) *Complex systems and archaeology.* University of Utah Press. [aHG]
Skinner, B. F. (1948) *Walden two.* Macmillan. [GA]
(1971) *Beyond freedom and dignity*. Knopf. [JIK]
Skyrms, B. (1996) *The evolution of the social contract.* Cambridge University Press. [KES]
Sloman, S. A. (1996) The empirical case for two systems of reasoning. *Psychological Bulletin* 119:3–22. [KES]
Slovic, P. (1995) The construction of preference. *American Psychologist* 50(5):364–71. [aHG]
Smith, A. (1759/1982) *The theory of moral sentiments*. Liberty Fund. [LA]
(1759/2000) *The theory of moral sentiments.* Prometheus. [aHG]
Smith, E. A. (2003) Human cooperation. Perspectives from behavioral ecology. In: *Genetic and cultural evolution of cooperation*, ed. P. Hammerstein, pp. 401–27. MIT Press. [RN]
Smith, E. A. & Winterhalder, B. (1992) *Evolutionary ecology and human behavior.* Aldine de Gruyter. [aHG]
Smith, V. (1982) Microeconomic systems as an experimental science. *American Economic Review* 72:923–55. [aHG]
Sober, E. & Wilson, D. S. (1998) *Unto others: The evolution and psychology of unselfish behavior*. Harvard University Press. [JWP]
Spinoza, B. (2005) *Ethics.* Penguin Classics. [rHG]
Stake, J. E. (2004) The property "instinct." *Philosophical Transactions of the Royal Society B: Biological Sciences* 359:1763–74. Available at: http://www.journals.royalsoc.ac.uk/(0i3l2rylficf2cbdgnwn442y)/app/home/contribution.asp?referrer=parent&backto=issue,12,16;journal,23,125;linkingpublicationresults,1:102022,1. [ODJ]
Stanovich, K. E. (1999) *Who is rational? Studies in individual differences in reasoning.* Erlbaum. [aHG, KES]
(2004) *The robot's rebellion: Finding meaning in the age of Darwin.* University of Chicago Press. [KES]
Stanovich, K. E. & West, R. F. (2000) Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences* 23:645–726. [KES]
Starmer, C. V. (2000) Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38:332–82. [DJZ]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press

Stein, E. (1996) *Without good reason: The rationality debate in philosophy and cognitive science.* Oxford University Press. [KES]
Stephens, W., McLinn, C. M. & Stevens, J. R. (2002) Discounting and reciprocity in an iterated prisoner's dilemma. *Science* 298:2216–18. [aHG]
Sternberg, R. J. & Wagner, R. K. (1999) *Readings in cognitive psychology.* Wadsworth. [aHG]
Stich, S. P. (1990) *The fragmentation of reason.* MIT Press. [KES]
Strauss, C. & Quinn, N. (1997) *A cognitive theory of cultural meaning, vol. 9.* Cambridge University Press. [DPK]
Striedter, G. F. (2005) *Principles of brain evolution*. Sinauer [RBG]
Sugden, R. (1993a) An axiomatic foundation for regret theory. *Journal of Economic Theory* 60(1):159–80. [aHG]
(1993b) Thinking as a team: Towards an explanation of nonselfish behaviour. *Social Philosophy and Policy* 10:69–89. [AMC]
(2005) The logic of team reasoning. In: *Teamwork: Multi-disciplinary perspectives*, ed. N. Gold, pp. 181–199. Palgrave Macmillan. [AMC]
Sugrue, L. P., Corrado, G. S. & Newsome, W. T. (2005) Choosing the greater of two goods: Neural currencies for valuation and decision making. *Nature Reviews Neuroscience* 6:363–75. [aHG]
Sutton, R. & Barto, A. G. (2000) *Reinforcement learning.* MIT Press. [aHG]
Taylor, P. & Jonker, L. (1978) Evolutionarily stable strategies and game dynamics. *Mathematical Biosciences* 40:145–56. [aHG]
Thayer, B. (2004) *Darwin and international relations*. University Press of Kentucky. [LA]
Tinbergen, N. (1963) On the aims and methods of ethology. *Zeitschrift fur Tierpsychologie* 20:410–33. [RLB, GEW]
Tomasello, M. (1999) The human adaptation for culture. *Annual Review of Anthropology*, 28:509–29. [DPK]
Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28(5):675–91. [aHG]
Tooby, J. & Cosmides, L. (1992) The psychological foundations of culture. In: *The adapted mind: Evolutionary psychology and the generation of culture*, ed. J. H. Barkow, L. Cosmides & J. Tooby, pp. 19–136. Oxford University Press. [aHG, MEP, JT]
Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology* 46:35–57. [aHG, MEP, GEW]
(1972) Parental investment and sexual selection. In: *Sexual selection and the descent of man, 1871-1971*, ed. B. Campbell, pp. 136–79. Aldine. [MEP]
(1974) Parent-offspring conflict. *American Zoologist* 14:249–64. [MEP]
(2000) The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences* 907:114–31. [MEP]
Tversky, A. & Kahneman, D. (1971) Belief in the law of small numbers. *Psychological Bulletin* 76:105–10. [aHG]
(1981) Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics* 106(4):1039–61. [aHG]
(1983) Extensional versus intuitive reasoning: The conjunction fallacy in probability judgement. *Psychological Review* 90:293–315. [aHG]
Tversky, A. & Slovic, P. & Kahneman, D. (1990) The causes of preference reversal. *American Economic Review* 80(1):204–17. [aHG, KRH]
Udéhn, L. (1992) The limits of economic imperialism. In: *Interfaces in economic and social analysis*, ed. U. Himmelstrand, pp. 239–80. Routledge. [GMH]
Uttal, W. R. (2001) *The new phrenology*. MIT Press. [ABM]
van den Berghe, P. L. (1990) Why most sociologists don't (and won't) think evolutionarily. *Sociological Forum* 5:173–85. [RLB]
Van Lange, P. A. M. (1999) The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology* 77:337–49. [AMC]
Vanberg, V. J. (2002) Rational choice versus program-based behavior: Alternative theoretical approaches and their relevance for the study of institutions. *Rationality and Society* 14(1):7–53. [GMH]
(2004) The rationality postulate in economics: Its ambiguity, its deficiency and its evolutionary alternative. *Journal of Economic Methodology* 11(1):1–29. [GMH, DJZ]
Velleman, J. D. (1992) What happens when somebody acts? *Mind* 101:461–81. [KES]
von Bertalanffy, L. (1968) *General system theory*. George Braziller. [RBG]
Von Neumann, J. & Morgenstern, O. (1944) *Theory of games and economic behavior.* Princeton University Press. [arHG]
Wason, P. C. (1966) Reasoning. In: *New horizons in psychology*, ed. B. Foss, pp. 135–51. Penguin. [aHG]
Wasserman, S. & Faust, K. (1994) *Social network analysis*. Cambridge University Press. [RAB]
Watts, D. J. (2002) A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences USA* 99:5766–71. [RAB]
Weisfeld, G. E. (1997) Discrete emotions theory with specific reference to pride and shame. In: *Uniting psychology and biology: Integrative perspectives on human development*, ed. N. L. Segal, G. E. Weisfeld & C. C. Weisfeld, pp. 419–43. American Psychological Association. [GEW]
Weiten, W. (2007) *Psychology: Themes and Variations, Seventh Edition*. Thomson Wadsworth. [RBG]
West, S. A., Gardner, A., Shuker, D. M., Reynolds, T., Burton-Chellow, M., Sykes, E. M., Guinnee, M. A. & Griffin, A. S. (2006) Cooperation and the scale of competition in humans. *Current Biology* 16:1103–06. [TG]
Westermarck, E. (1906) *The origin and development of the moral ideas*. Macmillan. [LA]
Wetherick, N. E. (1995) Reasoning and rationality: A critique of some experimental paradigms. *Theory and Psychology* 5(3):429–48. [aHG]
Whiten, A. (2001) Meta-representation and secondary representation. *Trends in Cognitive Sciences* 5:378. [KES]
Williams, G. C. (1966) *Adaptation and natural selection: A critique of some current evolutionary thought.* Princeton University Press. [aHG, JWP, MEP]
Williams, J. H. G., Whiten, A., Suddendorf, T. & Perrett, D. I. (2001) Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews* 25:287–95. [aHG]
Wilson, D. S. (2004) What is wrong with absolute individual fitness? *Trends in Ecology and Evolution* 19(5):245–48. [JWP]
Wilson, D. S. & Dugatkin, L. A. (1997) Group selection and assortative interactions. *American Naturalist* 149(2):336–51. [rHG]
Wilson, E. O. (1975) *Sociobiology: The new synthesis.* Harvard University Press. [rHG, TG, GEW]
(1998) *Consilience: The unity of knowledge.* Knopf. [aHG]
Wimsatt, W. C. (1980) Reductionistic research strategies and their biases in the units of selection controversy. In: *Scientific discovery, vol. 2: Historical and scientific case studies*, ed. T. Nickles, pp. 213–59. Reidel. [RBG]
Winter, S. G. (1971) Satisficing, selection and the innovating remnant. *Quarterly Journal of Economics* 85:237–61. [aHG]
Winterhalder, B. & Smith, E. A. (2000) Analyzing adaptive strategies: Human behavioral ecology at twenty-five. *Evolutionary Anthropology* 9:51–72. [RAB, EAS]
Wood, E. J. (2003) *Insurgent collective action and civil war in El Salvador.* Cambridge University Press. [aHG]
Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 6:111–78. [aHG]
Wrong, D. H. (1961) The oversocialized conception of man in modern sociology. *American Sociological Review* 26:183–93. [aHG]
Wylie, A. (1999) Rethinking unity as a "working hypothesis" for philosophy of science: How archaeologists exploit the disunities of science. *Perspectives on Science* 7:293–317. [SC]
Wynne-Edwards, V. C. (1962) *Animal dispersion in relation to social behaviour.* Oliver & Boyd. [JWP]
Young, H. P. (1998) *Individual strategy and social structure: An evolutionary theory of institutions.* Princeton University Press. [aHG]
Zajonc, R. B. (1980) Feeling and thinking: Preferences need no inferences. *American Psychologist* 35(2):151–75. [aHG]
(1984) On the primacy of affect. *American Psychologist* 39:117–23. [aHG]
Zizzo, D. J. (2002) Neurobiological measurements of cardinal utility: Hedonimeters or learning algorithms? *Social Choice and Welfare* 19:477–88. [DJZ]
(2003) Probability compounding in words and in practice. *Theory and Decision* 54:287–314. [DJZ]
(2005) Simple and compound lotteries: Experimental evidence and neural network modelling. In: *Transfer of knowledge in economic decision making*, ed. D.J. Zizzo. Palgrave Macmillan. [DJZ]

https://doi.org/10.1017/S0140525X0700088X Published online by Cambridge University Press