BEHAVIORAL AND BRAIN SCIENCES (1999) **22,** 923–989
*Printed in the United States of America*

# Color, consciousness, and the isomorphism constraint

**Stephen E. Palmer**
*Department of Psychology, University of California, Berkeley, Berkeley, CA 94720-1650*
**palmer@cogsci.berkeley.edu** **socrates.berkeley.edu/~plab**

**Abstract:** The relations among consciousness, brain, behavior, and scientific explanation are explored in the domain of color perception. Current scientific knowledge about color similarity, color composition, dimensional structure, unique colors, and color categories is used to assess Locke's "inverted spectrum argument" about the undetectability of color transformations. A symmetry analysis of color space shows that the literal interpretation of this argument – reversing the experience of a rainbow – would not work. Three other color-to-color transformations might work, however, depending on the relevance of certain color categories. The approach is then generalized to examine behavioral detection of arbitrary differences in color experiences, leading to the formulation of a principled distinction, called the "isomorphism constraint," between what can and cannot be determined about the nature of color experience by objective behavioral means. Finally, the prospects for achieving a biologically based explanation of color experience below the level of isomorphism are considered in light of the limitations of behavioral methods. Within-subject designs using biological interventions hold the greatest promise for scientific progress on consciousness, but objective knowledge of another person's experience appears impossible. The implications of these arguments for functionalism are discussed.

**Keywords:** basic color terms; color; consciousness; functionalism; inverted spectrum; isomorphism; qualia; subjectivity; symmetry

In this target article I discuss the relations among mind, brain, behavior, and science in the particular domain of color perception. My reasons for approaching these difficult issues from the perspective of color experience are twofold. First, there is a long philosophical tradition of debating the nature of internal experiences of color, dating from John Locke's (1690/1987) discussion of the so-called inverted spectrum argument. This intuitively compelling argument will be an important historical backdrop for much of this target article. Second, from a scientific standpoint, color is perhaps the most tractable, best understood aspect of mental life. It demonstrates better than any other topic how a mental phenomenon can be more fully understood by integrating knowledge from many different disciplines (Kay & McDaniel 1978; Palmer 1999; Thompson 1995). In this target article I turn once more to color for new insights into how conscious experience can be studied and understood scientifically.

I begin with a brief description of the inverted spectrum problem as posed in classical philosophical terms and then discuss how empirical constraints on the answer can be brought to bear in terms of the structure of human color experience as it is currently understood scientifically. This discussion ultimately leads to a principled distinction, called the *isomorphism constraint,* between what can and cannot be determined about the nature of experience by objective behavioral means. Finally, I consider the prospects for achieving a biologically based explanation of color experience, ending with some speculation about limitations on what science can achieve with respect to understanding color experience and other forms of consciousness.

## 1. Detecting transformations of color experience

The basic intuition that underlies Locke's (1690/1987) inverted spectrum argument is that other people might have the same overall set of color experiences you do but they might be differently connected to objects in the external world. When you and I look at the same red apple under the same lighting conditions, for example, do we have the same internal experience of redness, or might I have the experience you call greenness, or yellowness, or some other color? The issue is perhaps most clearly captured by the situation of looking at a rainbow. You perceive a particular ordering of chromatic experiences from red at the top to violet at the bottom, but I might perceive exactly the reverse of your experiences, with violet at the top and red at the bottom. We would both name the color at the top "red" and the one at the bottom "violet," of course, because that is what

STEPHEN E. PALMER is Professor of Psychology, Director of the Institute of Cognitive Studies, and Director of the undergraduate major in Cognitive Science at the University of California at Berkeley. He is the author of over 60 scientific publications in the area of visual perception, including his new textbook *Vision science: Photons to phenomenology* (1999, MIT Press), which provides an introduction to all aspects of vision from an interdisciplinary perspective. He is a member of the Society of Experimental Psychologists, a Fellow of both the American Psychological Association and the American Psychological Society, and a member of the Governing Board of the Psychonomic Society.

© 1999 Cambridge University Press 0140-525X/99 $12.50

https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

we have all been taught. Color naming is presumably mediated by internal color experiences, but it is only the sociolinguistically sanctioned stimulus-response associations that matter in our ability to communicate about colors. The fact that we name the same objects and lights with the same color terms is hence insufficient to determine whether our internal chromatic experiences are the same.

The rainbow-reversal interpretation of the inverted spectrum argument is quite literal in the sense that we have supposed that my experiences of the spectrally pure (monochromatic) colors are simply the "inverse" of yours about the spectral midpoint. Notice that the "inverted spectrum argument" is actually something of a misnomer: It is not the spectrum that is inverted – that is, nothing has happened to the rainbow itself – but the inner *experiences* in response to the spectrum. It would therefore be more accurate to call this the "inverted color argument." And because Locke's essential point is not limited to literal inversion, but could apply equally well to any possible mapping of color experiences in one observer (e.g., you) to the same set of color experiences in another observer (e.g., me), we will call it the "transformed color argument." The problem it poses is whether any such color-to-color transformation (excluding the identity mapping) could accurately describe the relation between our color experiences without the differences being objectively detectable. Because we do not have direct access to each other's internal experiences, the question boils down to whether any such color-to-color transformation could exist without being detectable through systematic differences in our behavior.

Notice that the transformed color argument as formulated above presupposes two important conditions: (1) that the two observers (canonically referred to as "you" and "me") *have* experiences in response to different spectra of electromagnetic energy, and (2) that their overall sets of color experiences are the same. Dropping one or both of these assumptions leads to different versions of the more general "color problem." Relaxing condition (2) suggests the possibility that you and I both have experiences in response to different light spectra but that one or both of us has at least some experiences the other does not. Our experiences might overlap to some degree, or they might even be completely disjoint, so that all my chromatic experiences in response to light spectra are qualitatively different from all of yours. Relaxing condition (1) suggests that I might have no experiences of color whatsoever in response to different light spectra and yet behave as you do with respect to them. I would be a "color zombie," able to name colors properly and produce standard data in various color discrimination and matching experiments, but without having any corresponding experiences at all. These possibilities will come to the fore later in this target article, but for now we will concentrate on the standard form of the problem in which it is assumed that both you and I have the same set of color experiences, whatever those might be, and ask whether they can be shown to be "differently arranged," so to speak.

The first problem is how to get a scientific handle on this philosophical problem of whether transformed color experiences could be detected in publicly observable behavior. Locke presumed that we could not, but there are many scientifically documented aspects of color-related behavior that bear critically on the answer. I will argue that whether there exist any undetectable color-to-color transformations can be recast into the simple question of whether an empirically accurate model of human color experience contains any symmetries.

**1.1. Color similarities.** In scientific discussions, color experience is usually described in terms of spatial models. Perhaps the simplest and best known is the color circle devised by Newton (1704), by now familiar to artists and much of the general public (see Fig. 1). It is an example of a *color space:* a multidimensional spatial representation (or model) in which different color experiences correspond to different points in the model. The locations of points representing colors are chosen so that the degree of psychological similarity between pairs of colors corresponds to the distance between the corresponding points in the model. In the color circle, the *spectral colors* of the rainbow are positioned in order along most of the circle's perimeter, and the *nonspectral colors,* including many reds, all magentas, and most purples, are located along the perimeter between the blue-violet and orange-red limits of the visible spectrum, as indicated in Figure 1 by the location of the color names outside the circle.

The color circle is a useful model of many aspects of color experience because it is easy to comprehend yet manages to capture an immense number of facts about the relations among color experiences in a highly economical fashion. The fact that red is perceived as more similar to orange than to green, for example, is reflected in the fact that the point representing red is closer to the point representing orange than it is to the point representing green. Corresponding

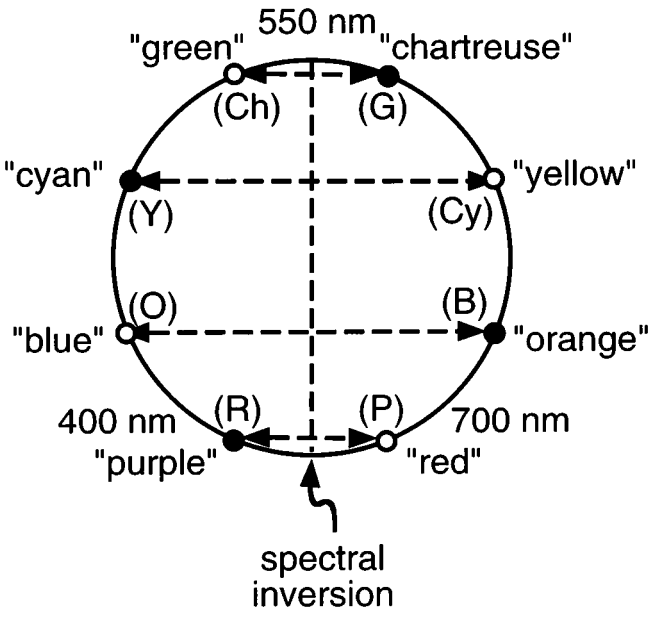



Figure 1. Newton's color circle and spectral inversion. Colors are arranged along the perimeter of a color circle, as indicated by the names on the outside of the circle. Open circles correspond to unique colors (red, green, blue, and yellow), which look subjectively pure. The dashed diameter indicates the axis of reflection corresponding to literal spectral inversion, and the dashed arrows indicate corresponding experiences under this transformation. Letters in parentheses inside the circle indicate the color experiences a spectrally inverted individual would have to the same physical stimuli a normal individual would experience as the colors indicated on the outside of the circle.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

similarity relations among triples of color experiences – that blue is more similar to purple than to yellow, and so on – are thus faithfully preserved in the distance relations among triples of points in the color circle. This correspondence between psychological similarity relations and spatial proximity relations lies at the heart of Shepard's (1962a; 1962b) elegant nonmetric method of multidimensional scaling, and recovering the color circle from similarity data was one of his first demonstrations of its use.

We will now employ Newton's (admittedly simplistic) color space to illustrate how such models relate to the inverted spectrum argument and why Locke's intuition seems initially so compelling. Consider once again the case of literal spectral inversion, which we will henceforth call "rainbow reversal" to be perfectly clear. The hypothesis that my color experiences are rainbow-reversed relative to yours would mean that my color circle is the reflection of yours about the dashed line shown in Figure 1. The abbreviations inside the circle indicate the color experiences I have in response to the same lights that you experience as indicated on the outside of the circle. Thus, your red corresponds to my violet (and vice versa), your orange to my blue (and vice versa), your yellow to my cyan (and vice versa), and your chartreuse to my green (and vice versa), as indicated by the dashed arrows perpendicular to the axis of reflection. The question is whether this color-to-color transformation could be detected by behavioral means.

It might seem at first blush that similarity judgments among colors would reveal rainbow reversal, but, in fact, they would not. You would say that orange is more similar to red than to green, but so would I, even though this would correspond internally to my experiencing blue as more similar to violet than to chartreuse. Indeed, we would make all the same relative similarity judgments about rainbow colors despite the enormous differences in our internal experiences of them. With respect to color similarity judgments among rainbow experiences, then, Locke was right: Rainbow reversal cannot be detected behaviorally from such data.

The reason rainbow reversal is not behaviorally detectable from such color similarity judgments is that the empirical model they specify (i.e, the basic color circle) is *symmetric* about the axis of reflection that corresponds to reversing the rainbow. A symmetry in a spatial model is a transformation that maps the model onto itself so that it is the same before and after the transformation. Rainbow reversal is thus a symmetry of the color circle – but so is any central reflection or rotation. Indeed, the only thing that makes any pair of such reflected or rotated color circles different is the nature of the internal experiences themselves. Because these are private events, any differences between yours and mine can be assessed only indirectly through our publicly observable behavior, as Wittgenstein (1953) argued so forcefully. The general claim is that any symmetry in a behaviorally constrained color space necessarily specifies a color-to-color transformation that cannot be detected by the behaviors that constrain the spatial model.

**1.2. Color composition.** There is much more that we know about color experience from behavioral observations, however, and these facts must also constrain the color space whose symmetries we seek to understand. One important factor is the *composition relations* among colors: how some color experiences can be analyzed into combinations of other, more basic color experiences. Most colors of the rainbow are binary in the sense that they appear to be composed of (or are analyzable into) two of the four primary chromatic colors: red, green, blue, and yellow.[1] Oranges, for example, seem to contain both redness and yellowness, purples seem to contain both blueness and redness, and so forth. In contrast, there are particular shades of red, green, blue, and yellow that do not appear to be composed of any other colors.[2] For example, there is a particular red, called *unique red*, that appears purely red, with no traces of yellowness, blueness, or greenness in it. There are similarly defined unique colors for green, blue, and yellow, each of which is pure in the same sense of having no traces of the other primary colors. Moreover, pairs of these primaries are related to each other as polar opposites: red versus green, and blue versus yellow. These important hypotheses about compositional relations among color experiences were pointed out by Ewald Hering (1878/1964), who used them as the basis of his opponent process theory of color vision.

Relations of color composition are not the same as or reducible to the relations of color similarity discussed above. As the color circle shows, the similarity relations among red, blue, and purple are essentially the same as those among orange, red, and purple (see Fig. 1). Even so, purple looks like it is composed of red and blue, whereas red does not look like it is composed of orange and purple. A complete model of color experience therefore requires that the uniqueness of the four chromatic primaries be represented within color space. It also requires a representation of the composition of colors in terms of how they are perceived to be analyzed into the four primaries.

Unique red, green, blue, and yellow can be represented by singular points at diametrically opposed locations in the color circle, and color composition relations of binary colors by their projections onto its orthogonal red – green and blue – yellow axes. Notice that this elaboration of the color circle now breaks many of the symmetries that were previously candidates for undetectable color-to-color transformations based only on color similarity relations. Only seven remain: the four central reflections about the dimensional axes and their angular bisectors and the three central rotations of 90°, 180°, and 270°. Rainbow reflection has been eliminated because, if my color experiences were related to yours by this transformation, I would judge purple, chartreuse, orange, and cyan to be unique colors rather than red, green, blue, and yellow. All other transformations that map unique colors to binary colors are likewise eliminated.

It is now clear how one might proceed in a scientific evaluation of the transformed color argument: Document evidence of the existence of asymmetries in a behaviorally constrained model of human color experience. If all symmetries can be shown to be broken, then Locke was wrong; my internal color experiences cannot be a transformation of yours without the difference being detectable in my behavior relative to yours. The focus must be on finding asymmetries rather than finding symmetries because the nature of scientific hypothesis testing requires ruling out null hypotheses (i.e., finding that differences are present) rather than accepting them (i.e., failing to find such differences). In the context of symmetries, this amounts to detecting asymmetries where there are differences rather than detecting symmetries where there are no differences.

**1.3. Asymmetries in lightness.** The color circle we have been using as a model of human color experience is inade-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

quate, however, primarily because it leaves out the vast majority of color experiences, including white, black, all their mixtures with each other (grays), and all their mixtures with the chromatic colors along the perimeter of the color circle. These color experiences are missing because the color circle is only two-dimensional, whereas the full set of color experiences is three-dimensional. Figure 2 shows a more complete spatial model of human color experiences that reflects color similarity relations in all three dimensions of perceived surface color and the full set of six compositional primary colors: red, green, blue, and yellow (as before) plus black and white.[3]

The experiences of surface colors represented in Figure 2 are classically defined by three dimensions, which we will call "hue," "saturation," and "lightness."[4] (For lights, the third dimension is usually designated as "brightness" rather than "lightness.") The dimension we normally associate with the basic color of a surface or light is called its "hue." In color space, hue corresponds to the angular direction in the horizontal plane from the central axis of color space to the location of the point representing that color. The second dimension of color space, called "saturation," represents the vividness of color experiences. It corresponds to the perpendicular distance from the central axis to the position of the color experience in color space. For example, the vivid colors of the rainbow lie along the outside edge because they have maximum saturation. All the grays lie along the central axis because they have zero saturation. The "muted," "muddy," and "pastel" colors in between have intermediate levels of saturation. The third dimension of surface color is called "lightness." In the coordinates of color space, lightness refers to the height of a color's position as it is drawn in Figure 2. All surface colors have some value on the lightness dimension, although it is perhaps most obvious for the achromatic grays that lie along the central axis, with white at the top and black at the bottom. In particular, it is worth noting that browns are represented as dark yellows and oranges (i.e., ones low in lightness), as indicated in Figure 2.

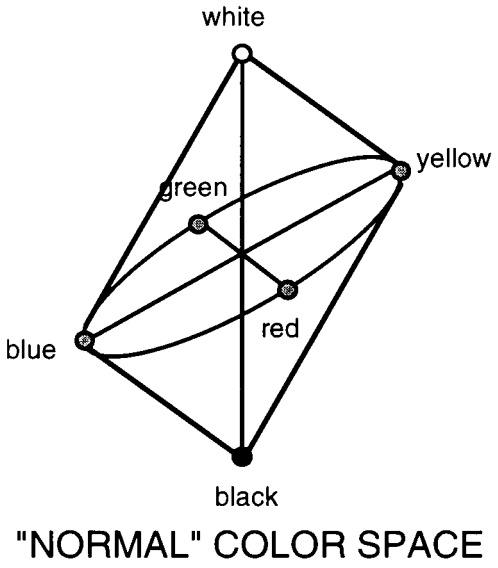



Figure 2. Three-dimensional color space. Colors are represented as points in a three-dimensional space according to the dimensions of hue, saturation, and lightness. The positions of the six unique colors (or Hering primaries) within this space are shown by circles.

The color circle corresponds to the perimeter of an oblique section through this color solid. This section is oblique because the most saturated yellows are quite light (and therefore higher in color space), whereas the most saturated blues and purples are quite dark (and therefore lower in color space), with the most saturated reds and greens at intermediate lightness values. This variation in lightness of maximum saturation colors breaks further symmetries in color space. A rotation of 180° about the vertical axis, for example, would map yellow to blue and blue to yellow. This color transformation could be detected behaviorally, however, because you would judge unique yellow to be lighter than unique blue, whereas I would judge the reverse. Because the color solid has no rotational symmetries, any simple rotation of color space about its lightness axis can be ruled out as a behaviorally undetectable color-to-color transformation. That is, rotations of three-dimensional color space can be detected and cannot be used to support the color-transformation argument.

There are still three approximate reflectional symmetries of the three-dimensional color spindle shown in Figure 3 that are likely to escape behavioral detection except in the most precise psychophysical tasks. One is reflection of red and green in the blue-yellow-black-white plane (Fig. 3A). This works because, at least to a first approximation, red and green are the same in lightness, and blue and yellow are mapped to themselves. A second symmetry is reversing both the blue-yellow and black-white axes (Fig. 3B). This solves the lightness problem with reversing blue and yellow because it also reverses the lightness (black-white) dimension. The third symmetry is the composition of the other two: namely, reversing all three axes, red for green, blue for yellow, and black for white (Fig. 3C).

Although all three of these transformations are logically possible, by far the most plausible is reflecting just the red-green dimension. Indeed, a persuasive argument can be made that such red-green-reversed perceivers may actually exist in the population of so-called "normal trichromats" (see Nida-Rümelin 1996). The argument goes like this: Normal trichromats have three different pigments in their three cone types. Some people, called *protanopes,* are red-green color blind because they have a gene that causes their long-wavelength (L) cones to have the same pigment as their medium-wavelength (M) cones. Other people, called *deuteranopes,* have a different form of red-green color blindness because they have a different gene that causes their M-cones to have the same pigment as their L-cones. In both cases, people with these genetic defects lose the ability to experience the red-green dimension of color space because the visual system codes this dimension by taking the difference between the outputs of the M- and L-cones. Now, suppose that someone had the genes for *both* forms of red-green color blindness simultaneously. Their L-cones would have the M-pigment, and their M-cones would have the L-pigment. Such people would, therefore, not be red-green color blind at all, but simply red-green-reversed trichromats.[5] They should exist. Assuming they do, they are proof that this color transformation is either undetectable or very difficult to detect by purely behavioral means, because no one has ever managed to identify such a person.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

**1.4. Basic color categories.** Harrison (1973) and Hardin (1997) have argued that there are further hurdles for transformed-color arguments to clear concerning the implications of color categories. Although it is not entirely clear that such categories reflect basic facts about color experience rather than some later cognitive stage of processing, but if they do, they have important consequences for the behavioral detectability of color transformations. To explain their import on color transformation arguments, we will have to review briefly some background claims about color naming and categorization.[6]

In their ground-breaking studies of cross-cultural color naming, Berlin and Kay (1969) found that there are a very small number of *basic color terms* (BCTs) across all the languages they studied. These BCTs refer to a corresponding set of *basic color categories* (BCCs) into which color experience can be partitioned. (We will assume the obvious one-to-one mapping between BCTs and BCCs in the discussion that follows.) Further research and analysis have postulated three different types of BCTs: *primary, derived,* and *composite* (Kay & McDaniel 1978). The most basic are the six primary categories: RED, GREEN, BLUE, YELLOW, BLACK, and WHITE. From these, six more categories are "derived" by the fuzzy-logical AND-ing (via fuzzy-set intersection; see Zadeh 1975) of two primary color categories:

GRAY = WHITE AND BLACK,
ORANGE = RED AND YELLOW,
PURPLE = RED AND BLUE,
BROWN = BLACK AND YELLOW,
PINK = WHITE AND RED,
GOLUBOI (a Russian word) = WHITE AND BLUE.[7]

Notice that this set does not include all possible combinations of primary BCCs. Some are ruled out by the structure of color space itself, such as red-green and blue-yellow, which cannot exist because they simply do not overlap and therefore have no exemplars in their fuzzy-logical intersection. Other combinations could exist as BCTs, such as blue-green, but do not for reasons that are as yet unknown.

The four "composite" color categories are formed by the fuzzy-logical OR-ing of two or more primary color categories:

WARM = RED OR YELLOW,
COOL = GREEN OR BLUE,
LIGHT-WARM = WHITE OR WARM = WHITE OR RED OR YELLOW,
DARK-COOL = BLACK OR COOL = BLACK OR GREEN OR BLUE.

Again, not all possible combinations of primary BCCs exist as composite BCTs. It seems reasonable that they be restricted to combinations of nearby primary BCCs in color space, ruling out RED OR GREEN and BLUE OR YELLOW. However, it is not clear why there are either no or few composite BCTs in known languages for RED OR BLUE, GREEN OR YELLOW, WHITE OR COOL, or BLACK OR WARM. These and other mysteries remain to be solved.

It is important to realize that these facts about BCTs are relevant to the present discussion only if they reveal important asymmetries in the structure of human color experiences. For example, if for some reason there are more just-noticeable-differences (jnds) between unique red and unique yellow than there are between unique green and unique blue, the wider psychophysical gap might explain why there are BCTs for ORANGE in many languages, but not for CYAN (blue-green). The fact that there are strong (possibly even universal) constraints on the BCTs that have been discovered in a large number of natural languages suggests that some basic neural mechanisms of human color vision are likely to be responsible. The most plausible alternative explanations are that the constraints on BCTs reflect structure in the nature of the environment (e.g., perhaps there are more salient orange-colored objects than cyan-

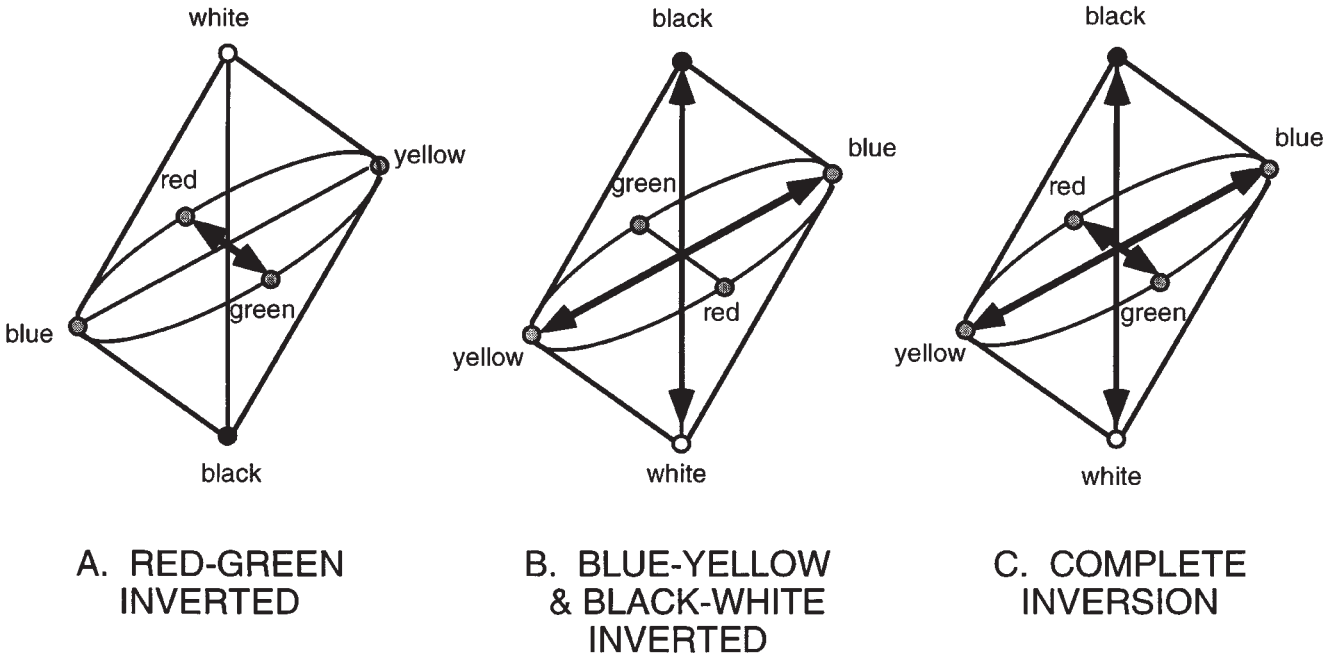



Figure 3. Approximate symmetries of color space. The color space depicted in Figure 2 has three approximate symmetries: reversal of the red-green dimension only (A), reversal of the blue-yellow and black-white dimensions (B), and reversal of all three dimensions (C).



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

colored objects), the nature of an organism's fit to its ecological niche (e.g., perhaps distinctions between different shades around orange are more important to the organism than those around cyan), or the patterns of contact and influence between languages.

If the empirical constraints on BCTs do, in fact, reflect underlying nonhomogeneities in the structure of color experience, they can be used to break further symmetries of color space. Primary BCTs, for which the evidence of universality is strongest, actually do not break any of the symmetries illustrated in Figure 3. This is because they correspond to the six Hering primaries (red, green, blue, yellow, black, and white) described above, which have already been taken into account by the unique points and axes of the color space.

Adding the derived and composite BCTs, however, breaks all further symmetries. Consider first derived BCTs. Red-green reversal (Fig. 3A) is ruled out, for example, by the asymmetry between the frequency of derived BCTs for ORANGE and PURPLE (frequently found) versus CYAN and CHARTREUSE (infrequently or never found). If I were red-green reversed – and if derived BCTs reflect intrinsic nonhomogeneities in my color experience – I should find it strange that BCTs were distributed in this way rather than in the opposite way. I should also find it odd that there is a BCT for PINK rather than light green. These facts pose no problem for blue-yellow and black-white reversal (Fig. 3B), however, because ORANGE maps to PURPLE (and vice versa) and PINK maps to itself.

Table 1 shows how different BCTs break the three candidate symmetries illustrated in Figure 3. The entries in this table were generated from Kay and McDaniel's (1978) analysis as follows. First, the transformation indicated at the top of each column dictates the remapping of primary BCCs as shown in the first six rows. Red-green reversal (column 1), for example, only requires changing RED to GREEN and GREEN to RED, where the first term indicates the original system of BCCs and the second designates the transformed system of BCCs. Next, the appropriate substitutions are made in Kay and McDaniel's formulas (see above) for the derived and composite BCCs. If the resulting formula for a new BCC corresponds to one for an old BCC, then a "+" is placed in the column for that transformation and the row of the original BCT. If not, then an "×" is entered there. To illustrate, consider the row for PURPLE. After red-green reversal (column 1 of Table 1), PURPLE = RED AND BLUE becomes PURPLE = GREEN AND BLUE. Because GREEN AND BLUE is not the formula for a BCT in Kay and McDaniel's theory, an "×" is indicated in the PURPLE row of the first column, meaning that this transformation maps this BCC into a non-BCC. After blue-yellow and black-white reversal (column 2 of Table 1), however, PURPLE = RED AND BLUE becomes PURPLE = RED AND YELLOW. Because RED AND YELLOW is the formula for a BCT (namely, ORANGE), a "+" is entered in the table, indicating that this transformation maps a BCC into another BCC. For complete reversal (column 3), PURPLE = RED AND BLUE becomes PURPLE = GREEN AND YELLOW, which is not a BCT; hence the "×" in column 3.

The validity of this analysis rests on the validity of Kay and McDaniel's original theory, of course, including the set of BCTs they enumerate and their definitions in terms of fuzzy-logical operations. If new BCTs have been discovered in the meantime or if new formulas have been proposed, the analysis of Table 1 will be correspondingly wrong in detail. However, the general nature of the reasoning is sound within this qualitative theoretical framework, and a more complete or accurate theory can be substituted for Kay and McDaniel's original one. Notice also that the analysis in Table 1 involves only approximate, qualitative evaluations of asymmetries, not metric evaluations. As long as there is a BCT in the general neighborhood of the transformed BCT, the relation is counted as symmetrical (+) in Table 1. As it turns out, metric precision is unnecessary because the symmetries are broken qualitatively.

The analysis in Table 1 shows that no color-to-color transformations survive a thorough BCT analysis intact, indicating that all symmetries are broken by the behavioral constraints implied by BCTs. I am not aware, however, of any behavioral data that directly support these asymmetries for derived and composite color categories in color experience. In many cases, the difference between derived or composite BCTs and non-BCTs is subtle enough that direct introspections are too blunt an instrument with which to decide. I myself would be hard-pressed to claim, for example, that it seems "better" or "more natural" to me that there is a BCT for light reds (PINK) than for light greens, independent of the fact that my language actually has a BCT for light reds and not for light greens. The case for ORANGE and PURPLE over blue-green and yellow-green seems somewhat more compelling. Even so, it would be hard to tell how much of such preferences for derived and composite BCCs over non-BCCs is the product of sociolinguistic training and how much the asymmetries in my underlying color experiences.

The existing evidence most relevant to these asymmetries comes from Rosch's (Heider's) studies of learning color terms in the Dani tribe of New Guinea (Heider 1972). In a classic cross-cultural experiment, Rosch found that the Dani, who have BCTs only for LIGHT-WARM and DARK-COOL, were able to learn new categories for RED, BLUE, GREEN, and YELLOW more easily than new categories for ORANGE, PURPLE, CYAN, and CHARTREUSE. This result shows that primary BCCs appear to be preferred over other color categories for the Dani – and presumably other cultures with composite BCTs – even though these categories are not overtly expressed in the BCTs of their language.

It is not yet clear whether this distinction would also be supported for derived or composite BCTs – which are the only ones that break the symmetries in Figure 3 – because Rosch's studies with the Dani examined the learnability only of primary BCTs. Some derived BCTs were used in the study (ORANGE and PURPLE), but they were actually employed in the contrasting non-BCT "control" categories. Moreover, her results, which have traditionally been interpreted in terms of the learnability of primary BCTs, can be explained equally well by color composition relations based on the four chromatic Hering primaries. The latter explanation has the advantage of a clearer basis in phenomenology and physiology than is available for BCTs in general.

The main question is whether the derived and composite BCTs are grounded firmly enough in color experience for the asymmetries they imply to be detected. Perhaps a new category for light green would be just as easy to learn as PINK for people whose language has neither BCT, and perhaps a new BCT for blue-green or yellow-green would


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

Table 1. *Basic color terms over three reflectional symmetries*[a]

| | Reflectional transformations | | |
|---|---|---|---|
| BCT (Kay–McDaniel) | R–G | B–Y/Bk–Wh | R–G/B–Y/Bk–Wh |
| RED (R) | + (G) | + (R) | + (G) |
| GREEN (G) | + (R) | + (G) | + (R) |
| BLUE (B) | + (B) | + (Y) | + (Y) |
| YELLOW (Y) | + (Y) | + (B) | + (B) |
| BLACK (Bk) | + (Bk) | + (Wh) | + (Wh) |
| WHITE (Wh) | + (Wh) | + (Bk) | + (Bk) |
| GRAY (Gr=Wh&Bk) | + (Wh&Bk=Gr) | + (Bk&Wh=Gr) | + (Bk&Wh=Gr) |
| PURPLE (P=R&B) | × (G&B= Ø) | + (R&Y =O) | × (G&Y= Ø) |
| ORANGE (O=R&Y) | × (G&Y= Ø) | + (R&B =P) | × (G&B= Ø) |
| BROWN (Br=Y&Bk) | + (Y&Bk=Br) | + (B&Wh=Gb) | + (B&Wh=Gb) |
| PINK (Pk=R&Wh) | × (G&Wh= Ø) | × (R&Bk= Ø) | × (G&Bk= Ø) |
| GOLUBOI (Gb=B&Wh) | + (B&Wh=Gb) | + (Y&Bk=Br) | + (Y&Bk=Br) |
| WARM (Wm=RvY) | × (GvY= Ø) | × (RvB= Ø) | + (GvB=C) |
| COOL (C=GvB) | × (RvB= Ø) | × (GvY= Ø) | + (RvY=Wm) |
| LT-WARM (LW=WhvRvY) | × (WhvGvY=O) | × (BkvRvB= Ø) | + (BkvGvB=DC) |
| DK-COOL (DC=BkvGvB) | × (BkvRvB= Ø) | × (WhvGvY= Ø) | + (WhvRvY=LW) |

[a]& = fuzzy logical AND; v = fuzzy logical OR; Ø = no corresponding BCT; + = symmetric BCT present; × = no symmetric BCT present.

be just as easy to learn as ORANGE or PURPLE. The strongest argument for a phenomenologically privileged status of a derived BCC can be made for BROWN, insofar as it seems qualitatively different from the yellow and orange hue families of which it is part (Hardin 1997).

The crucial question at the center of this issue is whether the structure associated with BCCs is caused by nonhomogeneities of color experience. The most obvious way to document such effects of categories on perceptual experience is to look for so-called categorical perception phenomena. "Categorical perception" refers to a phenomenon in which small changes in certain stimulus continua *across* a categorical boundary produce large changes in perceptual experience, whereas corresponding changes *within* category boundaries produce much smaller changes in experience. The classic case is the effect of continuous acoustical variables on categorical perception of phonemes (e.g., Liberman et al. 1957).

In the color domain, the evidence is mixed. On the one hand, categorical effects in color perception have been reported by several researchers, even in infants (e.g., Bornstein et al. 1976) and monkeys (e.g., Sandell et al. 1979), for whom linguistic labels cannot be the mediators of such effects. On the other hand, these categorical effects are seldom as sharply defined as for categories of speech perception, and the fuzziness and ineffability of category boundaries is well documented in many other studies (e.g., Berlin & Kay 1969; Rosch 1973). Moreover, the categorical effects that have been reported are typically restricted to the primary BCTs of red, green, blue, and yellow, leaving us, once again, with an open question about the status of derived BCTs.

One phenomenon of normal experience that can be viewed as supporting categorical structure in color experience is the banded appearance of the rainbow (Hardin 1997). The physical continuum of photon wavelength that underlies the rainbow is purely quantitative and unidimensional, with no physical divisions that would produce "bands" of any sort. Why, then, does a wavelength rainbow appear banded? One possibility is that qualitative distinctions between color categories are directly represented in perceptual experience, as Hardin (1997) has argued, and that these produce qualitatively distinct bands in the appearance of the rainbow.

There is an alternative to the categorical explanation that must be considered, however. Because all chromatic colors (except the four unique ones) are experienced as mixtures of different amounts of red, green, blue, and yellow, the banded appearance of the rainbow might arise simply from the gradual transitions between these qualitatively different colors. In this case, the bands are attributable to color composition rather than to color categories. The most obvious way to discriminate between these two possibilities is to ask whether orange is perceived as a distinct band, qualitatively different from the adjacent reds and yellows, whether it is perceived merely as a transition between them, or whether it is something in between. The BCC view predicts a separate orange band because of the existence of the derived color category for orange, whereas the compositional view predicts no such band. If people do experience a separate orange band, there is the further question of whether this band is present only in the perceptions of people who speak languages with a BCT for ORANGE or whether it appears universally. Unfortunately, we do not yet have the answers to these deceptively simple questions.

Whether derived and composite BCTs are grounded in color experience may seem like a fine point, but, as Table 1 shows, it has crucial implications for the possibility of be-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

haviorally detecting color transformations. If BCCs are not reflected in color experience, or if only primary BCCs are reflected, then the prior conclusion stands that there are at least three transformations of color space that may well escape behavioral detection. If composite and/or derived BCCs are relevant, then no form of the color transformation argument will actually work.

**1.5. Asymmetries in color similarity.** Thus far, we have been assuming that all aspects of color experience can be naturally represented within a spatial model such as the one shown in Figure 2, but this is not necessarily true. One potential problem concerns systematic asymmetries in color-similarity relations. An axiomatic property of all metric dimensional spaces is that distances between points are symmetrical: The distance from A to B is the same as that from B to A (Krantz et al. 1971). If spatial distance is to represent experienced (dis)similarity, then color similarity relations must also be symmetrical.

Rosch (1975) has reported similarity results that contradict this assumption with respect to focal versus nonfocal colors (which correspond approximately to unique and binary colors, respectively). When Rosch had subjects indicate the perceived similarity between one color (the target) and another (the standard), she found small but systematic effects: Nonfocal targets were perceived as more similar to focal standards than vice versa (e.g., off reds were judged to be more similar to true red than true red was to off reds). Although these effects were not large, they are noteworthy for at least two reasons. One is that they create a serious problem for capturing all relations among colors in a purely spatial model. The other is that they may constitute another kind of evidence that color categories influence color experience.

Even so, Rosch's results do not necessarily rule out the possibility that certain color transformations can escape behavioral detection. There are two issues. The first concerns how these asymmetries in similarity are distributed in color space. The focal colors for the primary BCCs are essentially the unique primaries, and we have already noted that the three transformations of Figure 3 preserve their uniqueness. Asymmetrical distortions in distance relations with respect to these unique points can also be preserved by certain transformations, as illustrated in Figure 4. This diagram shows a hypothetical representation of asymmetries in similarity relations within the color circle that would be preserved by the same reflections that preserve the four unique points. The magnitude and direction of the asymmetries are represented by vectors indicating the directional difference in similarity between the four primary focal colors (large circles) and various comparison colors (small circles with arrows). This diagram shows that such asymmetries in similarity could be symmetrically distributed in a color space with respect to the primary focal colors. We simply do not have enough information on this issue to arrive at a firm conclusion.

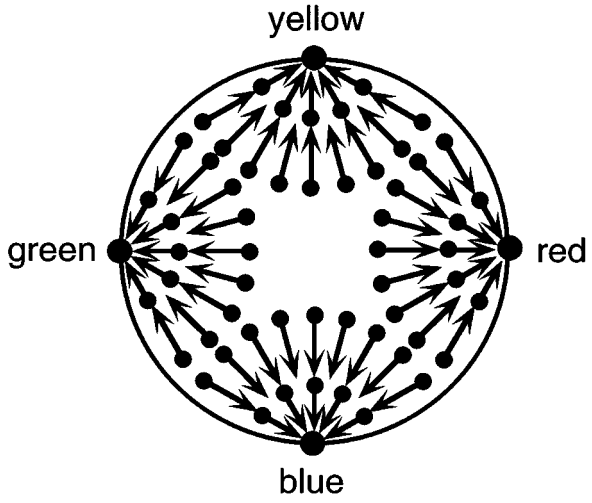



Figure 4. A symmetrical distribution of distance asymmetries. Rosch's (1975) results show that nonfocal colors are judged to be more similar to focal colors than the latter is to the former. The arrows represent the degree and direction of asymmetry between the four primary focal colors (large circles) and nonfocal colors (small circles with arrows attached).

The second critical issue for the behavioral detectability of color transformations is whether asymmetrical similarities to focal colors hold just for primary categories or whether they also hold for derived categories. As we have already noted, derived categories break all global symmetries of color space, so asymmetries in similarity with respect to these focal colors (e.g., ORANGE, PURPLE, PINK, BROWN, and GOLUBOI) would allow detection of any color-to-color transformations. On this point, I know of no evidence. Rosch found asymmetries in color similarity using only the primary focal colors – RED, GREEN, BLUE, and YELLOW – but did not test for asymmetries in derived color categories. The primary focal colors are symmetrically distributed and thus may cause no problems for any of the candidate symmetries of color space in Figure 3. A further question is whether these asymmetries are actually caused by color categories or color composition, which has a clearer and more obvious bearing on color experience. Again, we do not yet know the answer.

**1.6. Metrical asymmetries.** There are other potential sources of asymmetry in color space that might reflect nonhomogeneities in color experience that could be detected behaviorally. One concerns the metrical structure of color space as measured by the discriminability of color samples. Suppose, for example, that unique red is perceived to be more different from unique blue than unique blue is from unique green. Given that they are all unique versions of chromatic primaries, it seems plausible that they are, in some sense, equally different, and this is the rationale for placing them at opposite poles of orthogonal diameters of the color circle in Figure 1. But there are other ways of determining distances in color space psychophysically, such as counting the number of jnds between pairs of colors. Each jnd is measured psychophysically by finding the difference threshold: the smallest difference along a continuum that can just barely be detected. Using this method, red and blue might well prove to be more different from each other than blue and green. Munsell color space, which is based on measurements of equally spaced differences in hue, represents this difference in discriminability by a greater distance between red and blue, and MacLaury (1997a) has reported data supporting the same conclusion. Other metrical differences might also prove to be asymmetric when measured by counting jnds, and, if they are reliably different, they break what otherwise might be plausible symmetries.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

Let us summarize our discussion about the possibility of detecting color transformations empirically. There are just three candidate transformations that survive the most basic behavioral constraints concerning color experience as reflected in the global structure of color space: red-green reversal, blue-yellow and black-white reversal, and reversal of all three axes (red-green, blue-yellow, and black-white), as illustrated in Figure 3. If the color space depicted in Figure 2 is at least roughly accurate, then all three transformations will preserve the similarity relations among colors, the dimensional description of colors in terms of hue, saturation, and lightness, the decomposition of colors into the six Hering primaries, and the distinguished points of the six unique colors.

Adding color categories and color-naming data into the mix makes the situation more complex. The same three transformations survive further constraints owing to the primary BCCs and BCTs, including the distribution of the six primary color categories in color space and the asymmetries in similarity relations around the focal colors for the four chromatic primary color categories. Composite and derived BCCs rule out all transformations, however, by breaking their symmetries, but only if they are the result of intrinsic properties of the color system (i.e., based on experiential factors) rather than to extrinsic ecological factors (i.e., based on the physical environment or sociolinguistic community).

I have argued that the crucial issue in assessing the validity of the transformed-color argument is the existence of symmetries in an empirically accurate model of color experiences. Because no such model presently exists, the exercise is premature for reaching firm conclusions. I have used color spaces as the focus of this enterprise because they are by far the dominant modeling tool for this domain and because the nature of global transformations is particularly transparent within them. The argument from symmetry is not limited to spatial models, however; it can be applied to neural-network models, abstract-propositional models, or any other sort of model. The only requirements are that the set of possible color-to-color transformations can be specified in the model and that the results of such transformations can be assessed in terms of the requisite empirical constraints. If the behavior of the model is invariant over the transformation, it is symmetric with respect to that transformation, and the transformed-color argument will work.

## 2. The isomorphism constraint

The questions to which I now turn concern which aspects of mental life scientists can hope to study and understand objectively and which we cannot, using color experiences as the example. In the present section I will discuss the limitations of behavioral science, and in section 3, I will consider the possibility that biological science can take us beyond these limits.

**2.1. The subjectivity barrier.** It is universally agreed that there is a behaviorally defined subjectivity barrier with respect to how much others can know about our experiences, and color experiences are no exception. Some aspects of experience can be shared across observers, whereas others cannot be. We know that many aspects of color experience must be shared across observers because normal trichromats agree in their linguistic statements and other sorts of discriminative behavior with respect to colors. Color-blind individuals also agree with others having the same form of color deficiency, but they do not agree across color-deficiency classes or with normal trichromats. These aspects of color experience are therefore objectively shared and fully available to behavioral science. Other aspects are indeterminate in this respect, however, in that they appear to be free to vary without affecting any known aspect of behavior. They are purely subjective and therefore unavailable to behavioral science. Even if they happen to be identical across observers, scientists would never know with certainty that this was the case. In this section I attempt to define this barrier between objective and subjective phenomena with respect to behavioral science.

I suggest that the two relevant aspects of experience are the intrinsic qualities of experiences themselves versus the relational structure that holds among those experiences. These two aspects are normally so completely intertwined that it may seem perverse to advocate separating them, but if they lie on different sides of the subjectivity barrier, as I suggest, then it is important to make the distinction.

The most obvious aspect of visual awareness is certainly the nature of the experiences themselves, such as the sensory quality of redness or circularity, to pick two examples at random. It seems that the quality of these experiences is flat-out impossible to define behaviorally, given that we have access to no one's experiences but our own. This is why color-to-color transformations are a legitimate problem in the first place: The quality of individual experiences lies beyond the behavioral subjectivity barrier.

One might suppose that there is at least one aspect of experiences that can be specified behaviorally, namely, their *individuality.* The set of colors that a person can individuate (discriminate), for example, determines whether someone has full trichromatic color experience or a restricted set owing to some form of color blindness. Notice, however, that experiences can be individuated behaviorally only by asking people to discriminate between two stimuli, responding "same" or "different" to various pairs. Color-blind individuals reveal their reduced set of color experiences by performing at chance in discriminating between certain color samples that normal individuals distinguish quite easily. Thus, even individuating experiences behaviorally is actually about the *relation* between two (or more) experiences by designating whether they are the same or different.

**2.2. The importance of relational structure.** The second aspect of experience is one to which behavioral science does have access: namely, structure among experiences carried by their relations to each other. Regardless of what the experience of red itself is like, normal trichromats agree that it is more like orange than green. Likewise, sighted observers agree that a circle looks more like a regular octagon than an equilateral triangle. These relations among experiences are just as important as the qualities of experiences per se – and in certain respects, even more so – because they determine the *structure* of experience, which can be shared despite the subjectivity barrier. If experiences had no relational structure, they would simply be a collection of completely different and totally unrelated mental states, like the "blooming buzzing confusion" that William James suggested is the nature of sensory experience in infants (James 1890/1950). Without relational structure, for ex-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

ample, we would not experience colors as a coherent domain of experience, more similar to each other than they are to shapes: Redness would be as much like circularity as it is like greenness – or middle-C, or the smell of freshly ground coffee, or the taste of pumpkin pie.

Relational structure is even more crucial within a single domain of experience such as color. Without it we would not experience white as being lighter than gray or gray as being lighter than black; we would experience them only as different colors, and equally different at that. Because of lighter-than relations among color experiences, we are aware of the ordering of colors in terms of the continuous dimension of lightness, ranging from black to white. Indeed, the entire structure of color space (see Fig. 2) is determined by relations among colors, particularly relations of composition and similarity. These relations provide the rich, complex dimensional superstructure of color experience. It is quite literally unimaginable what color experience would be like without this structure.

I have argued at some length that it may not be possible to be sure that my experiences of colors are the same as your experiences of colors strictly from our behavior, because mine could just as easily be some structure-preserving transformation of yours. We can (and do) agree on basic color terms that refer to them – I call roses "red," violets "blue," and so forth – so that we can communicate effectively about individual colors. This is an objective behavioral fact about color experience, but it tells us absolutely nothing about the quality of those experiences except that they are discriminably different from others. It is an exceedingly weak constraint in the same sense that a nominal scale is the weakest type of measurement system (Stevens 1951). We merely learn to attach the same labels to our corresponding internal experiences that arise from viewing the same collections of wavelengths, regardless of what our particular internal experiences might be.

Further constraints are introduced, however, once we begin to consider binary or higher-order relations among experiences (Krantz et al. 1971). Both you and I can make judgments about the relative similarity of two colors to a third, or the relative lightness of two colors, for example. These inherently relational judgments are also objective in that normal trichromats agree about them, at least within some margin of error. This is not to say that my relational experiences are the same as yours or that we can even determine whether or not they are. My experiences of color similarity relations might be as wildly different from yours as my individual color experiences are from yours, but the *structure* of our experiences and relations can nevertheless be identical.

Preserving relational structure appears to be a necessary condition for one set of objects to represent another (Palmer 1978). Indeed, model theory formalizes the situation in which one set of objects models another in terms of the existence of a function that maps objects in the first set to objects in the second set, so that corresponding relations are preserved in a precise, set-theoretic sense (Tarski 1954).[8] This requirement explains why the same three-dimensional color solid is able to model color experiences for all normal trichromats, even if they happen to have wildly different color experiences: It captures exactly the relational structure among the color experiences for each individual by mapping them onto points in space so that relations among color experiences are preserved by spatial relations among points in the color solid (see Fig. 5). This is not to say that my experience of white is literally *above* my experience of black, even though the point representing white in the color solid is above the one representing black in the conventional color solid. However, my experience of white is *lighter than* that of black, and the relation *above* in canonical color space corresponds to the relation *lighter than* in color experience and preserves its structure.

The emerging picture is that the nature of color experiences cannot be uniquely fixed by objective behavioral means, but their structural interrelations can be. This means that, logically speaking, *any* set of underlying experiences will do for color, provided the experiences relate to each other in the required way. The same argument can be extended quite generally to other perceptual and conceptual domains, although both the underlying experiential components and their relational structure will be different. The experience of musical pitch, for example, could be grounded in any of an infinite variety of experiential dimensions, but it would always have to have the same double-helical structure characteristic of perceived pitch relations (Shepard 1982). Although these are both cases in which there is a clear geometrical structure associated with the experiential domain, this need not be true. The only requirement is that there be some kind of relations among the experiences that constitute their structure.

**2.3. Symmetry, isomorphism, and relational structure.** In section 1, I argued at some length that the existence of symmetries in color space is the key issue in assessing the validity of color-transformation arguments. I now return to this topic to ask why this might be the case and how the answer relates to the above discussion of the structure of experiences.

Mathematically, symmetries are functions that have two special properties, known as "automorphism" and "isomor-

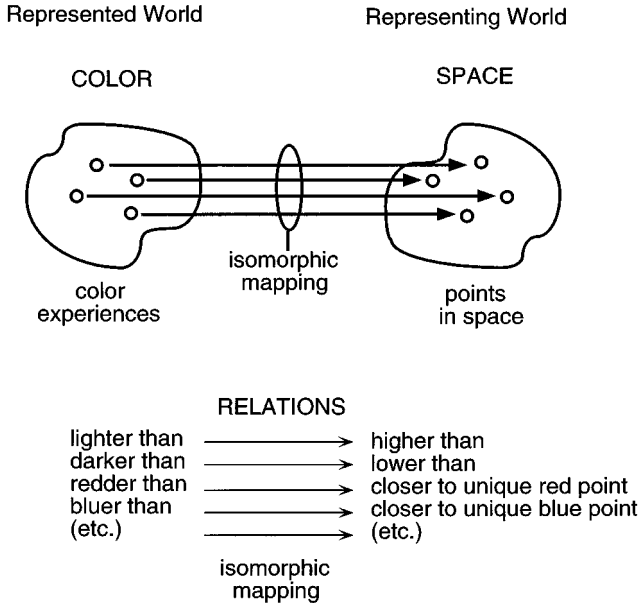



Figure 5. The color/space isomorphism. Color experiences are mapped to points in a multidimensional space (see Fig. 2) such that color relations (e.g., *lighter-than, redder-than*) are preserved by corresponding spatial relations (e.g., *higher-than, closer-to-unique-red-point*).


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

phism." They are *automorphisms* because they map a given domain onto itself in a one-to-one fashion. In the case of symmetries in color space, the automorphic function therefore maps a color space onto itself. Automorphism is critical for Locke's (1690/1987) argument because, as was mentioned at the outset, it assumes that the two observers in question have the same overall set of color experiences. The only question is whether their color experiences might have different causal connections to the outside world.

The second special property of symmetries is that they are *isomorphisms.* "Isomorphism," which means roughly "having the same structure," is a mathematical function in which, intuitively speaking, one set of entities is mapped onto another set of entities such that the structure of relations among the first set is preserved by the structure of corresponding relations among the second set (e.g., Tarski 1954).[9] Figure 5 illustrates the general requirements for an isomorphism to hold using the (presumed) isomorphic relation between color experience and color space as an example. The function maps color experiences onto points in a dimensional color space such that relations among color experiences (*lighter than, more similar than,* etc.) are preserved by corresponding relations among corresponding points in space (*higher than, closer to,* etc.). This allows a direct and valid translation from facts about the relations among color experiences into facts about the relations among points in color space. The only thing missing is the capability to say anything about the nature of the color experiences themselves (or the nature of the points themselves) except that they satisfy corresponding relations. The fact that such an isomorphism can be constructed is the principal reason that spatial models are good representations of color experiences. Notice further that, because isomorphism is both transitive and symmetric, if your color experience and my color experience are isomorphic to the same color space, then our color experiences are necessarily isomorphic to each other.[10]

Isomorphism is crucial for the transformed-color argument because, I suggest, the only kinds of differences that can be detected behaviorally are differences in relational structure, and relational structure is precisely what is preserved by isomorphism. I can say (or otherwise indicate behaviorally) that color A is lighter than color B, but I cannot in any way communicate how light either one appears to me in absolute terms. I can also say with confidence that red is more similar to orange than it is to green, but I cannot express either similarity in absolute terms. It might seem that one can make "absolute" ratings of, say, the lightness of individual color samples on a rating scale, but such ratings are, in fact, always relative to the range of possibilities.

Both automorphism and isomorphism are required to satisfy the assumptions of Locke's (1690/1987) original inverted spectrum argument. There are other versions of the more general color question that do not require the automorphism component, however. If you and I both have color experiences, for example, but they are not the same overall set, then automorphism does not hold. The nonoverlap can vary from minimal to complete. It would be minimal, for example, if everything seemed just a bit lighter to me than to you. In this case, no new dimensions are involved, just different values over an expanded range of lightness. My experience of pure white would then be lighter than any experience you have ever had, and your experience of pure black would be darker than any experience I have ever had. This mapping between our experiences is not automorphic because there are some color experiences – my white and your black – unique to each of us.

More radically, though, my color experiences might be totally different from your color experiences in ways neither you nor I can imagine. This would be the case if my experiences of the three dimensions of color space were qualitatively different from yours, as though we lived in completely different subspaces of some hypothetical "experiential hyperspace." I would have chromatic dimensions for what we all call red-green and blue-yellow variations, just as you would, but they would span hue dimensions qualitatively different from any you have ever experienced. The existence of such additional experiential dimensions of color can be inferred from comparative studies of color vision, which show that some animals have four or even five dimensions of color experience (see Thompson et al. 1992). At least some of the dimensions of chromatic experiences of such animals, whatever they may be, must be qualitatively different from any of yours or mine. There is certainly no logical requirement that my experiences of the range of hues (or saturations or lightnesses) be anything like your experiences of them. Biological considerations can be brought to bear, as we will discuss shortly at some length, but there are enough differences between the brains of different perceivers to undermine an a priori assumption that the color experiences they give rise to are necessarily the same, except in extraordinary circumstances such as exact clones.

Such considerations lead me to conclude that automorphism is not central to understanding the general question of behavioral detectability of differences in color experiences, even though it is required for Locke's inverted spectrum argument. Isomorphism, however, appears to be key in evaluating the detectability of any color transformations under any circumstances. If your color experiences are isomorphic to mine, your experiences will be undetectably different from mine, because the structure in the relations among your color experiences is the same as that in the corresponding relations among my color experiences. And if only relations can be assessed by behavioral means, then isomorphism is the strongest form of equality that can be claimed for color experiences across observers based on behavior. The relations that must be structurally preserved include (at least) color similarity, color composition, unique versus binary colors, and dimensional ordering.

I will call this condition of structural equality to the level of isomorphism *the isomorphism constraint* and will suggest that it constitutes a fundamental limitation on what can be discovered about experience by behavioral methods. It means that, even if all the dimensions of my color experiences are qualitatively different from yours, we can still behave identically with respect to colors as long as our experiences are isomorphic.[11] My experiences would clearly have to be three-dimensional; would have to include six unique reference experiences (for the unique colors) at the poles of three axes; would have to include an angular dimension for hue, a radial dimension for saturation, and a linear dimension for lightness, and so forth. If all the relevant conditions were met, then my color experiences could be arbitrarily different from yours without the differences being behaviorally detectable.

Again, it is important to understand that none of these conclusions depends on there being spatial models of the cognitive domain. Experiential domains in which there are



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

viable spatial models make good illustrations because the idea of isomorphism is particularly clear in such cases, but spatial models are by no means necessary. The internal relations among experiences could even be fundamentally incompatible with spatial representations, conforming to none of the fundamental axioms of metric dimensional spaces. The only requirement is that, whatever the qualitative nature of your internal experiences and the relations among them may be, the relations among my corresponding experiences must have the same structure.

If the shared aspects of experience do indeed coincide exactly with structural relations – that is, what is preserved by an isomorphism – the argument thus far can be summarized as follows: *Objective behavioral methods can determine the nature of experiences up to, but not beyond, the criterion of isomorphism.* The subjectivity barrier would then coincide precisely with the isomorphism constraint.

It is interesting that this criterion of isomorphism is not unique to the subjectivity barrier or to behavioral science; it also exists in axiomatic formulations of mathematics. In classical mathematics, a domain is formalized by specifying a set of primitive elements (e.g., points, lines, planes, and three-dimensional spaces in geometry) and a set of axioms that specify the relations among them (e.g., two points uniquely determine a line, three noncollinear points a plane, etc.). Given a set of primitive elements, a set of axioms, and the rules of mathematical inference, mathematicians can prove theorems that specify many further relations among mathematical objects in the domain. These theorems are guaranteed to be true if the axioms are true.

However, the elements to which all the axioms and theorems refer cannot be fixed in any way except by the nature of the relations among them; they refer equally to any entities that satisfy the set of axioms. That is why mathematicians sometimes discover that there is an alternative interpretation of the primitive elements, called a "dual system," in which all the same statements hold. For example, the points, lines, planes, and spaces of projective geometry in three dimensions can be reinterpreted as spaces, planes, lines, and points, respectively, because all the same relations hold when the elements in the latter system are substituted systematically for their corresponding dual elements in the former system. All the same axioms hold; therefore, all the same theorems are true. An axiomatic mathematical system can, therefore, be conceived as a complex structure of mathematical relations on an underlying, but otherwise undefined, set of primitives that are free to vary in any way. As Poincaré (1952) observed: "Mathematicians do not study objects, but the relations between objects; to them it is a matter of indifference if these objects are replaced by others, provided that the relations do not change" (p. 20). The same can be said about behavioral scientists with respect to consciousness: We do not study experiences, but the relations among experiences. It is (or should be) a matter of indifference to behavioral scientists if the experiences of one person differ from those of another, provided that the relations among experiences are the same.

**2.4. Relation to functionalism.** The analysis we have just given of color experience bears important relations to certain aspects of functionalism. One salient characterization of functionalist accounts of the mind is that they are based on the causal relations among mental states and their input and output relations to the external world (Dennett 1978; Fodor 1968; Putnam 1960). Two cognitive systems that have the same causal relational structure (i.e., that have corresponding causal relations among all their corresponding mental states) and the same causal relations to the external world (on both the input and the output ends) are functionally equivalent. Functionalist doctrine claims that such systems, which we can call "causally isomorphic" (but not causally equivalent), are therefore mentally equivalent.

However, the analysis in section 1 suggests that this is not necessarily so. In particular, two systems having the same causal relational structure (including input-output relations to the environment) can have radically different conscious states. There are at least two ways in which this can happen. One was discussed at length in section 1, namely, the possibility that the experiences of two people may be quite different even though they have the same relational structure, as in the behavioral undetectability of causally isomorphic color experiences. The other is that one of the systems might have the same causal relational structure over its internal states and their relations to the world, yet have no conscious experiences whatever. Let us now consider this latter case more carefully.

Suppose we were to create a working "color machine" that actually processes information from light in the same way that people do and that responds as people typically do. This is a reasonable goal. Figure 6 illustrates one way to construct the "front end" of such a machine. It analyzes incoming light using prisms, cardboard masks, photometers, electronic adding and subtracting circuits, and so forth to process color information according to the principles of color perception as they are currently understood. The details of how we now believe receptor outputs are integrated to compute the dimensional values of color space may be wrong, of course, but the crucial issue is whether substituting the right computations would result in the machine having color experiences. For such a machine to respond behaviorally to light patches, it would have to be extended by adding processes to produce basic color terms for the colors it is shown, to analyze the composition of colors into their compositions in terms of the Hering primaries, to make color-similarity ratings, and so forth. Moreover, it would have to do all this in a way that is behaviorally and computationally equivalent to the way in which people perform these tasks. Supposing that such a machine could be constructed – and it would not be very difficult to do – it seems almost bizarre to claim that, because it derived the correct coordinates in color space for, say, unique red, named it "red," judged it more like orange than green, agreed that it was a "warm" rather than a "cool" color, and so on, it necessarily *had* an experience of intense redness. Rather, the machine appears to *simulate* color experiences without actually *having* them. This difference between having and simulating experience underlies Searle's (1980) distinction between "weak AI" and "strong AI."

Even so, it is surprisingly difficult to prove that this machine fails to have color experiences. A card-carrying functionalist would claim that such a machine does have color experiences purely by virtue of the computations it performs. That may seem unlikely to readers not in the grip of functionalism, but can it be refuted? The underlying difficulty is the "problem of other minds." Because we do not have access to the experiences of any other entity – be it a person, animal, or machine – how can we tell whether the


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

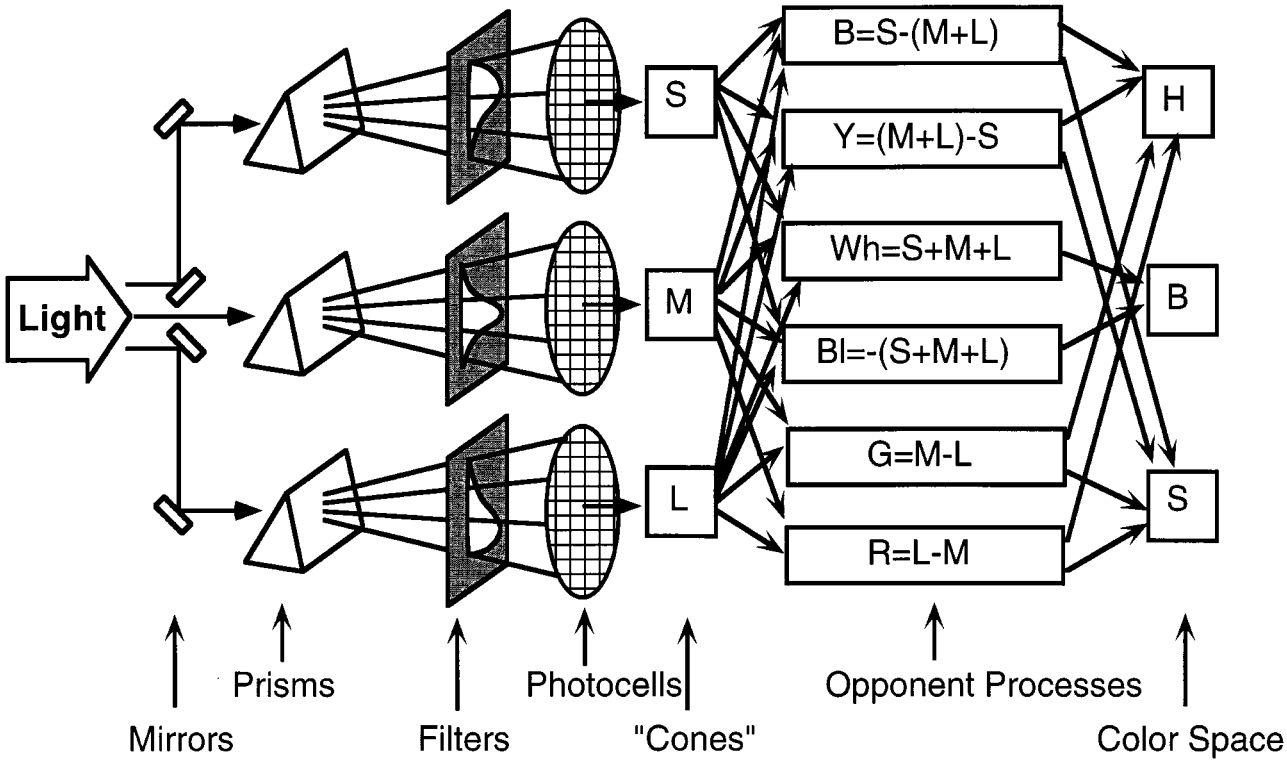



Figure 6. A color machine. A system of mirrors, prisms, cardboard masks, photocells, and computational circuits can derive values corresponding to the coordinates of colors in color space.

color machine actually has color experiences as a result of performing these computations? There seems to be a logical possibility that it might, but Searle (1980) has argued that it cannot. Let us therefore consider an adaptation of his argument to the present topic of color experience.

**2.5. The color room.** Searle's original thought experiment, known as the Chinese room argument, was not concerned with color perception but with understanding foreign languages. He asked whether a computer could literally understand Chinese if it were programmed so that when it was asked questions in Chinese, it answered in a way that was indistinguishable from native speakers of Chinese. The claim of "strong AI" is that such a computer actually understands Chinese, including whatever conscious experiences go along with such understanding, but Searle has claimed it does not.

Searle's argument can be adapted to make a corresponding, and in some ways simpler and more compelling, point about the insufficiency of performing information-processing operations to produce color experiences. Consider a system that takes as input the quantum catches of the three different types of receptors responsible for color vision (the cones sensitive to short, medium, and long wavelengths) and gives as output the name or description of the color in some language. The claim of strong AI is that such a system must be *having* color experiences rather than just *simulating* them. This is a form of functionalism, because the claim is that if the machine's internal color states and processes are causally isomorphic to people's color states and processes, then the machine literally has the analogous mental states, including any experiential component.

The thought experiment is to imagine yourself inside a "color room." You receive an ordered set of three numbers through the "input" door. You then consult a rule book that tells you how to perform the various numerical calculations necessary to arrive at three other numbers (corresponding to hue, saturation, and brightness of a light, although you do not know that this is what they represent). Ultimately, the rules in the book tell you how to use these values to select one of a particular set of letter strings to hand out the output door. If the language of color naming is German, and if you do not know German, these would be meaningless strings of letters to you: "schwarz," "weiss," "rot," "grün," "blau," and so on. You sit in the color room, carrying out your computations flawlessly, but totally mechanically, because nothing you are doing has any meaning, other than the fact that you are performing various arithmetical operations and following certain logical rules that govern which letter strings you pass through the "output" door.

Unbeknownst to you, however, people outside the room interpret the three numbers you receive as inputs to be the quantum catches of the three types of color-sensitive receptors in the retina in response to a colored light. They also interpret the meaningless (to you) letter strings you send out as basic color terms in German. Thanks to the rules in the book, the letter strings you send out in response to the three numbers turn out to be the appropriate German names for the color patches that produced the three input numbers.

Let us suppose that you become so well practiced at your task that your performance is indistinguishable from that of German speakers with normal color vision. Let us also suppose, for the sake of argument, that the rules you are fol-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

lowing in the book conform precisely to the established operations the visual nervous system uses to arrive at internal representations of colors and color names. These two facts mean that you have satisfied the usual functionalist criteria for claiming that you have mental states associated with color perception and naming. The key question is whether carrying out these computations would necessarily cause you to have the color experiences you normally would have when viewing the corresponding patch of color.

Most people's intuitions on this question are perfectly clear. Performing these computations, no matter how closely they might correspond to visual information processing in response to patches of color, would not give rise to anything like the experiences of color that would spontaneously arise if you were actually looking at the patch of light.

The force of this argument is that it avoids the problem of other minds in deciding whether a color machine has color experiences. By inserting yourself into the color room system of rules and computations, you know that there is a conscious agent within it who could have color experiences if that were indeed a necessary consequence of doing the right computations. Defenders of strong AI have attempted to refute Searle's (1980) argument on a number of grounds. We will not consider these arguments here; they are fairly involved and readily available in the commentaries following Searle's original article, together with his responses to them. Their translation into the color room version is straightforward.

It appears that behavioral methods are deficient, not only in determining the qualities of internal experiences but also in determining whether there are any internal experiences at all. They can determine the relational structure among experiences, but not their subjective quality, including whether they are simulated (i.e., nonexistent) or real. This appears to be a great deal less than one expects from a full scientific understanding of consciousness.

## 3. Biology to the rescue?

Having considered the inherent limitations of behavioral approaches to understanding color awareness, let us now consider the prospects for going further via biological approaches, as suggested by neuroscientists (e.g., Crick 1994). Will physiology succeed in penetrating the subjectivity barrier? In particular, will it allow us to go beyond the isomorphism constraint to get a handle on the quality of color experiences within and across individuals? If so, how? If not, why not?

**3.1. Correlational versus causal theories.** In considering the status of physiological hypotheses about experience, it is important to distinguish two different sorts, which I will call "correlational" and "causal" claims. Correlational claims concern the type of brain activity that takes place when experiences are occurring and fails to take place when they are not. Claims that conscious experiences arise only when there is brain activity at a particular location, within synapses employing a particular neurotransmitter, or firing at a particular rate would all be correlational, because they provide no causal explanation of how this particular kind of brain activity produces experience. They fail to fill the "explanatory gap" between ordinary physical events and experience (Levine 1983), because they merely designate a subset of neural activity with which consciousness is associated. No explanation is given for this association; it simply is the sort of activity that accompanies consciousness.

At this point it would be appropriate to contrast such correlational claims with a good example of a causal one: a theory that provides a scientifically plausible explanation of how a particular form of brain activity actually produces experience.[12] Unfortunately, no examples of such theories are at hand. In fact, to this writer's knowledge, no one has ever suggested any theory that the scientific community regards as giving even a remotely plausible causal account of how experience arises from neural events. This does not mean that such a theory is impossible in principle, only that we have yet to find a serious candidate.

A related difference between correlational and causal biological theories of experience is that they are likely to differ in generalizability. Correlations are inherently specific to the particular biological system within which they have been identified. In the best-case scenario, a good correlational claim about the neural substrate of human consciousness might generalize to chimpanzees and certain monkeys, possibly even to dogs or rats, but probably not to frogs or snails simply because their brains are too different. If a correlational analysis were to show that human consciousness is associated with, say, gammabutyric acid (GABA) activity, would that necessarily mean that creatures without GABA are not conscious? Or might some other evolutionarily related neural transmitter serve the same function in brains without GABA? If so, what features might they have in common? Even more drastically, what about extraterrestrial beings, whose whole physical makeup might be radically different from our own? In such cases, a correlational analysis is almost bound to break down. A causal theory of consciousness might have a fighting chance, however, because the structure of the theory itself could provide the lines along which reasonable generalizations might flow.

**3.2. The DNA analogy.** It seems conceivable that we will someday attain a full enough understanding of the brain that the biological mechanisms underlying conscious experience will be discovered and understood, even though we currently have few clues regarding what they might be like. The best we can do right now is appeal to an analogy to the understanding of the nature of life that was achieved by discovering the molecular structure of DNA.[13]

The facts to be explained by a theory of life concern certain differences in macroscopic properties of living organisms versus nonliving objects, particularly their abilities to grow, sustain their internal functions, and reproduce. The basic mechanisms available for a truly causal theory to explain such life functions are the physical behavior of ordinary matter. Prior to the discovery of the structure of DNA by Crick and Watson (1953), however, there was a huge explanatory gap between the known physical laws and these molar properties of living organisms. The gap was large enough that some theorists claimed there must be a special "vital force" that pervades living tissue and endows it with the requisite properties. Other theorists rejected this view, of course, insisting on a mechanistic theory devoid of mysterious, nonphysical forces.

In retrospect, the facts about living organisms that turned out to be most important for discovering the mechanisms of life concern their ability to reproduce. Especially


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

pertinent were the lawful inheritance relations between characteristics of offspring and characteristics of their parents. The general nature of these regularities was worked out by the brilliant Austrian monk, Gregor Mendel, in his painstaking experiments with garden peas. His results enabled him to propose, test, and refine a theory of inheritance based on hypothetical physical entities, now called "genes," that were combined in reproduction and expressed in offspring according to lawful principles. He worked out this genetic theory without ever directly observing the physical basis of these genes. Mendel's work is a beautiful example of how a molar behavioral analysis can give insights into underlying molecular mechanisms without those mechanisms being directly observed.

Mendel's genetic theory did not fill the explanatory gap, however, because the physical mechanism was not yet specified. Many years later, when sufficiently powerful microscopes enabled biologists to observe the internal machinery of cells directly, correlational hypotheses correctly suggested that the nucleus of cells was crucial in their self-replicating abilities. More refined correlational conjectures focused on the strands of chromosomes within the nucleus as the crucial structures involved in cell reproduction. Even so, the explanatory gap remained, because there were no serious candidate theories about how the properties of living cells might arise from chromosomes, whose internal structure was unknown.

The gap-filling discovery was made by Crick and Watson (1953) when they determined the double-helical structure of DNA. It revealed how a complex molecule, which was made of more basic building blocks, could unravel and replicate itself in a purely mechanical way. As further implications of this structure began to be worked out, it became increasingly clear how purely physical processes could form the basis of not only genetics, but cellular replication, protein synthesis, and a multitude of other unique and previously inexplicable properties of living tissue. The current biological understanding of life is thus a good example of how a truly causal theory can fill a seemingly immense explanatory gap in science. It did not appear magically, but was historically preceded by molar analyses of organismic phenomena and by correlational observations of the underlying biological mechanisms. Indeed, if Crick and Watson had worked out the structure of DNA prior to these other discoveries, its importance would very likely have gone completely unnoticed. Thus, all levels of theorizing were important in the historical development of a scientific theory of life.

In this analogy, our present knowledge of the biological substrates of conscious experience is best thought of as pre-Mendelian. After decades of actively ignoring the problem, scientists of many persuasions are beginning to work on it. Behavioral scientists are looking for functional correlates of consciousness in the hope of discerning regularities that would allow the formulation of computational theories (e.g., Baars 1988; Johnson-Laird 1983; Marcel 1983). Such theories can be thought of as akin to Mendelian genetics: hypotheses about the nature of consciousness at an abstract information-processing level distinct from the actual physical mechanisms that produce it.

Biological scientists are also beginning to look seriously for neural correlates of consciousness in the hope of narrowing the problem to some relevant subset of the brain where conscious experiences arise (e.g., Crick 1994; Crick & Koch 1995; Sheinberg & Logothetis 1997). This work can be likened to microscopic studies that identified the special importance of the nucleus and particularly of the strands of chromosomes in cellular division and reproduction. This biological analysis of the correlates of consciousness is barely in its infancy, and much work will be required before we will know where to look for the relevant mechanisms. And we are still very much in the pre-DNA phase of our quest to provide a causal explanation of conscious experience. No one has a clue as yet about what such a theory might look like. If and when it is discovered, it will be a scientific breakthrough of staggering importance and implications, for it will unlock one of the deepest scientific problems of all time. Many scientists believe this to be possible, perhaps even likely (e.g., Crick 1994). It may not turn out to be a purely reductionist account, as the explanation of genetics by DNA turned out to be, but most believe that there is a biological explanation and that it will eventually be found.

However, even if this is true, it is still unclear whether biological science will be able to take us beyond the limitations of behavioral science in understanding the nature of experiences themselves. By analogy with the DNA story, surely the first task is to find out what aspects of brain activity correlate with different experiences. Then we can ask whether this will tell us more about underlying experiences than we can find out behaviorally. Finally, we can ask about the prospects for achieving a truly causal explanation.

**3.3. Neural correlates of color experience.** Many questions about the neural correlates of color experience can be asked and answered within a biological framework. Some already have been, using present knowledge and technology. We know, for example, that the neural mechanisms underlying color experience must be somewhere in the cortex rather than in the retina or precortical visual system. One line of evidence is the existence of a form of color blindness or color weakness, called "achromatopsia," which is caused by damage to a certain region of the visual cortex between the occipital and parietal lobes of the brain (Meadows 1974). We do not yet know whether this region is actually the neural locus of color experiences or whether they occur farther along in the chain of neural processing, but we are slowly working our way toward this goal. It may involve activity in the frontal lobes, as Crick and Koch (1995) have suggested for consciousness in general, activity in the anterior cingulate gyrus, as Posner and Raichle (1994) have proposed, or activity in some other, as-yet-unsuspected structure in the brain. In any case, there seems little reason to doubt that someday the region(s) of the human brain in which color experiences arise will be discovered and that the pertinent properties of neural activity there will be identified. The question I now want to address is whether such discoveries are likely to tell us any more about color experience than we already know from our own experiences and from the results of behavioral studies. In asking this question, there is no reason for our inquiries to be restricted by the technologies presently available, so from here on we will resort to thought experiments in which we are freed from such limitations.

Let us first consider how correlations between brain events and color experiences could be studied and what might be discovered by doing so. Suppose I am a subject in an experiment in which the activity of every cell in my brain can be monitored by some futuristic "cerebrometer." Neu-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

roscientists could then study the differences between various forms of conscious and unconscious activity in my brain by correlating this neural activity with my color experiences. They could find out, for example, what particular patterns of neural activity in what particular regions of my brain correspond to my experiences of particular shades of red, orange, green, or any other color.

Presumably, the structure of these patterns of neural firings would turn out to be isomorphic to the structure of my color experiences and also to their representations in some suitably structured color space. This would mean, to take just one example, that the pattern of neural activity corresponding to red experience, whatever it might be, should be more like the pattern corresponding to orange experience than to the pattern corresponding to green experience. This is just the relation envisioned by Gestalt psychologists in their doctrine of psychophysiological isomorphism (Köhler 1929). I see no conceptual problems in carrying out this program of research, other than inventing the cerebrometer, of course. The results would tell us a great deal about the neural correlates of color experiences, but would they allow neuroscientists to determine the quality of my color experiences or even to determine whether my color experiences are the same as yours?

Notice that, although these correlations are intended to study relations between my brain's neural activity and my qualitative experiences, they cannot be computed from my experiences themselves, because I alone have access to these experiences. Rather, they would have to be computed from records of my *behavioral reports* of my experiences, because calculating conditional dependencies between two phenomena (in this case, patterns of brain activity and color experiences) requires repeated, objective, quantitative measurements of both. Because my experience can be objectively (and incompletely) assessed only through my behavior, the correlations will actually reflect dependencies between my brain activity and my behavior. That being the case, it is unclear how such a correlational approach will be able to go beyond the limitations of behavioral science in establishing the nature of color experiences, because behavioral methods are inherent within it.

This is not to say that we will fail to learn anything about the biology of color experience from the cerebrometer experiments. On the contrary, we will be able to learn everything there is to know about the biological correlates of experience up to the level of isomorphism. The results would tell us, for example, what the neural correlates are for hue, saturation, and lightness, or for redness/greenness, blueness/yellowness, and blackness/whiteness.[14] However, the correlated brain activity cannot inform us about the quality of the experiences themselves; these might still be any set of experiences that have the required relational structure. We have hit the isomorphism constraint again.

**3.4. Defining neurological equivalence classes of color experience.** What about the more modest goal of using biological methods to settle the color-transformation problem? Even if scientists cannot determine the exact quality of my (or your) color experiences from their neural correlates, can they at least determine whether our experiences of colors are the same? This is a weaker problem; it requires determining only whether you and I belong to the same equivalence class of color experiencers rather than determining what our experiences are actually like.

The obvious approach to this issue is simply to have neuroscientists, armed with their cerebrometers, correlate your brain activity with your reports of color experiences under controlled conditions, just as they did for mine, and then compare the two patterns of activity. Certain relations between our patterns of brain activity might superficially suggest certain specific kinds of color transformations. For example, less extreme firing rates in opponent process color cells in your brain might suggest a contraction of color space, or the reversal of firing rates in such cells of your brain might suggest an inversion transformation. However, only if your brain activity were *identical* to my brain activity in response to the same stimuli would one be justified in concluding that our experiences of the colors were necessarily the same. If they were identical, any other conclusion, however logically valid, would be eliminated by Ockham's razor as unparsimonious, for in the absence of any difference in brain activity it is simpler to suppose that our experiences are the same than that they differ in some mystical, ineffable way. This conclusion is consistent with the philosophical notion of supervenience (Kim 1984).

Assessing whether our patterns of brain activity are the same is not as straightforward as it might seem, however; it requires specifying a principled physical correspondence between our two brains. If those portions of our brains that are involved in color perception were identical, establishing this correspondence would not seem to pose much of a problem.[15] This situation might arise for a tiny subset of individuals, such as clones and identical twins, but the fact is that most people's brains differ from each other in a multitude of ways. Even seemingly minor physiological differences in color-relevant neurons, such as variations in baseline firing rates, the fine temporal structure of neural spiking, or the conduction speed of axons, might conceivably cause nontrivial differences in experience. In the absence of a causal theory of experience, it would be impossible to know which differences matter and which do not.

Even this problem of neural variation might not prove insurmountable if there were some way of studying the relation between biological differences and experiential differences. That would enable scientists to determine whether variations across individuals in, say, measures of baseline rates, fine temporal firing pattern, or axon conduction speed systematically influence experience. Here we run into the subjectivity barrier again, however, because assessing experiential differences between individuals requires somehow comparing their color experiences. As we have seen, such comparisons below the level of isomorphism are defeated by the inherent subjectivity of experience and its underdetermination by purely behavioral methods.

If there are objective differences in the relational structure of our experiences, such as occur in the various kinds of color blindness, appropriate behavioral methods can clearly detect them, and those differences can then be correlated with differences in brain activity. This enables us to determine a set of behaviorally defined equivalence classes of color perceivers – normal trichromats, protanopes, deuteranopes, and so on (see Fig. 7) – according to the relational color discriminations they can make. If our relational structures are identical, however, then our experiences are still free to vary within the isomorphism constraint without detection, and, if there are any potentially relevant differences between our brains that might produce experiential differences, it is unjustified to assume


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

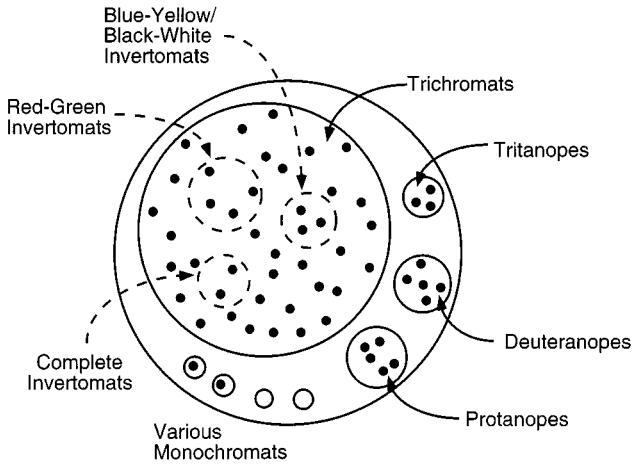



Figure 7. Equivalence classes of color perceivers. The Venn diagrams with solid lines indicate the behaviorally defined equivalence classes of color perceivers who have isomorphic color experiences, but not necessarily equivalent color experiences. Dashed circles indicate the possible existence of three classes of trichromats with color experiences that differ from those of normal trichromats at a subisomorphic level, corresponding to the three symmetries of color space indicated in Figure 3.

equivalence of color experiences. Exactly how one decides what kinds of differences between neural events are "potentially relevant" to experience is a serious problem, but one to which there are potential solutions, at least in principle (see below). Many factors will presumably be judged irrelevant by these criteria, but others are likely to remain.

It is important to note that even if we were able to conclude that two individuals' experiences are the same because their underlying physiology is the same, this would not specify the quality of their experiences to an outside observer. It would indicate only that, whatever they might be, they are the same. Only if I am one of the people whose experiences have been determined to be the same (or, from your perspective, if you are one of them) will this fact determine for me (or for you) what the experiences of another person are actually like. In other words, valid criteria of experiential sameness based on sameness of physiology would allow one to establish equivalence classes of experiential qualities, but it would determine the quality of those experiences only for the particular class to which the self belongs. An important question remains, however: How are we to determine which physiological differences matter for color experience?

**3.5. Within- versus between-subject designs.** We encounter the subjectivity barrier every time we try to make comparisons of subisomorphic differences across individuals. It therefore seems clear that progress in identifying experiential differences below the level of isomorphism can be made only by studying the problem *within individuals.* If pharmacologists were able to produce color-altering drugs, for example, their experiential and physiological effects could be studied and correlated within individuals. If administering "inverticillin," for example, produced a red-green reversal of color experiences within individuals, then some combination of its various physiological effects must have produced this change. Scientists would have to rule out certain bizarre, but logically possible, changes in language or memory as alternative explanations, but we will suppose that such alternative explanations can be effectively eliminated.[16]

It seems intuitively plausible that subisomorphic changes in experience owing to physiological interventions would be detected in appropriately designed within-subject experiments. We naturally suppose that, if the experienced world changed color over a period of seconds, minutes, or even hours, we would notice that this had happened and in what way the world had changed. Surely one would realize that grass now looks red instead of green, and stop signs green instead of red. Even so, various philosophers have tried to imagine scenarios in which such changes would not be detectable within an individual (e.g., Putnam 1965), but the conditions required are complex, convoluted, and generally unconvincing (e.g., Dennett 1991). The difficulty, in a nutshell, is that, although changing the color experiences would be (conceptually, though not practically) rather simple, the changes could easily be detected through mismatches with memories, such as noticing the oddity of seeing red grass. This is precisely the kind of change in experience that would allow subjects in our hypothetical experiment to report that the world looked different after the physiological intervention.

In fact, there is clear evidence that people can detect changes in color experience under circumstances similar to those we are suggesting. Strokes that damage particular areas of visual cortex (in the occipitoparietal area) result in spontaneous reports by patients that their color experience has disappeared or has lessened noticeably in intensity (Meadows 1974). Reports from these patients with achromatopsia thus provide evidence that changes in color experience can be detected within individuals by this means of memory comparisons in at least some cases. There is also evidence that other types of color changes go experientially undetected, however. Hardin (1990) has argued that because the lens yellows with age, the precise colors a person experiences will change, albeit glacially, during their lifetime, a change that no one seems to notice. The latter example is less relevant to our present concerns, however, because the changes in color experience we are contemplating are both swift and enormous compared to the slow, slight yellowing of the lens. Still, it is worthwhile to realize that we are making some assumptions in supposing that people will be able to report changes in color experience over time as a result of a biological intervention.

One important use of within-subject experiments is to determine whether certain physiological differences are relevant to color experience. If color experiences do not change when an intervention alters some aspect of an individual's neural structure or function, then that particular aspect can be ruled out as responsible for differences in color experience, at least in that individual. The situation is actually a bit more complex than this, because, even if a given factor does not influence color experience, it might have some influence in conjunction with another factor. The set of potentially relevant variables is thus the orthogonal combination of all possible single factors, which is presumably a very large set. In any case, each factor or combination of factors that can be eliminated enlarges the number of individuals in each neural equivalence class and thus increases the number of individuals who can be inferred to share a particular set of color experiences at the subisomorphic level.[17]

One of the most interesting findings about the subjective



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

quality of color experiences comes from a within-subject design, albeit one caused by nature rather than scientific intervention. The question addressed is how colors appear to color-blind individuals versus people with normal color vision. The subjectivity barrier thwarts direct, between-subject comparisons of experiences, even though we can obtain good behavioral evidence that people differ in the particular pairs of colors they can discriminate. But a within-subject comparison is possible for individuals who have one color-normal eye and one color-blind eye (MacLeod & Lennie 1976; Sloan & Wollach 1948). Although such people are extraordinarily rare, they report that some colors in their color-blind eye look essentially the same as in their normal eye, that others appear uncolored (gray) instead of colored, and that still others appear to be mixtures of grays and colors. Through a red-green color-blind eye, for example, such subjects report seeing everything in various shades of blue, yellow, and gray, as if the normal color solid were projected onto the blue-yellow-black-white plane of color space. This result is not definitive for the more general, between-subjects case because the two eyes of such a subject are connected to the same brain, and this may strongly constrain color appearances arising from stimulation in the color-blind eye. Even so, it is an interesting and important finding, one that is possible only because nature has provided an unusual within-subject design.

Many of the fascinating phenomena responsible for the recent surge of interest in consciousness can also be cast within this within-subject framework. One of the classic examples is blindsight, in which a biological intervention (surgical removal of occipital cortex in one hemisphere) caused an individual to fail to have experiences of the sort he used to have in one-half of his visual field (Weiskrantz 1986). Knowing full well what it was like to have visual experiences in his left visual field before the operation and what it was still like to have them in his right visual field, D.B. enabled scientists to conclude that something about the surgery had eliminated visual experience and gave impetus to the search to find out what was responsible. Unilateral neglect and other syndromes of organic brain damage tap into the same format of within-subject designs with biological interventions induced by natural causes. They are among the most powerful tools in current scientific methodology for exploring biological substrates of experience.

**3.6. Causal theories of color experience.** Thus far, we have considered only correlational studies of experience and argued that, unless the color systems within two brains are identical so that their experiential content can be assumed to be the same, comparisons between color experiences across individuals will be thwarted by limitations inherent in behavioral science. But what about a causal theory of consciousness? Might not such a theory allow between-subject comparisons of color experiences to be made directly? Might not such a theory even be able to determine the quality of different people's experiences of color in an objective manner?

Because we have no idea what a causal theory of consciousness might be like, it is impossible to determine whether it will contain the sort of mechanisms from which one could predict the specific quality of color experiences within and/or across individuals. For the sake of argument, though, let us suppose it purports to do so. A nontrivial further requirement of scientific theories, however, is that they be testable. Could a causal theory of the qualities of experience be tested?

Again the answer depends on whether the kinds of variations it predicts can be induced within individuals or only across individuals. If a physiological intervention – be it a drug, surgical procedure, or brain-controlling machine – can be identified that will reliably alter the crucial neural properties in the required ways, then it could be administered to individuals and its effects assessed by having subjects compare their color experiences before and after the intervention. If colors became systematically lighter, or if redness/greenness reversed, or if colors began to be experienced as sounds, as predicted by some causal theory, then it would be supported; otherwise it would not. Although such experiments might be technically difficult, they are conceptually straightforward.

If the required changes could not be induced within individuals, but only examined across them, we run into the subjectivity barrier again. Differences that can be measured behaviorally, which I suggest reflect differences in relational structure, can of course be detected and compared to the predictions of the theory. However, if the theory makes predictions that certain physiological differences will cause experiential differences across individuals below the level of isomorphism, they cannot be objectively tested. Biological methods can verify that the appropriate difference is present across individuals, of course, but behavioral methods will not be able to detect its subisomorphic effects on experience for the reasons we have already discussed.

The advantages of within-subject manipulations in revealing subisomorphic variations in color experience actually fit rather well with the intuitions on which the inverted-spectrum argument is based. The reason inverting the spectrum seems so compelling is that I can easily imagine what it would be like for me to experience such a transformation of color experience. If I awoke one morning to find my color experiences transformed, I believe I would know immediately that it had happened and how it was different. This is essentially the within-subject design I have argued is the only method one might be able to use to break the subjectivity barrier. However, this would not actually "break" it as much as merely avoid it because, within an individual, there *is* no subjectivity barrier.[18]

On the other hand, even within-subject experiments have serious limitations arising from the subjectivity barrier. They may enable scientists to detect experiential differences within individuals over time that cannot be detected across individuals, but they still cannot fix the quality of particular experiences. The examples we have discussed allow scientists to conclude only that the experiences *changed* in the particular way described by the subject, not how they were before the change or how they are after it. That is, an external observer can find out only about the *relation* between the experiences before and after the intervention. This does not allow any inferences about the actual qualities of either set of experiences. For example, if some drug produced a red/green reversal in a subject, that fact could be determined, but it would not constrain the nature of either green or red experiences in any way. This difficulty should sound familiar; it is the isomorphism constraint again. It arises because the subjectivity barrier is still in place, not between one subject and another, but between the subject and the experimenter. It will always be there as long as scientific methods require objective measurement.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

## 4. Implications for functionalism

The theory of mind most relevant to the considerations raised in this target article is functionalism because of the central role relations among mental states play in both cases. In particular, functionalism identifies the nature of mind with the relations among mental states and their causal connections to the environment. Crucial to the functionalist view is the assumption that the nature of mental states does not depend on their particular physical realizations. Minds must be physically embodied *somehow* to have the necessary causal connections to the world, of course, but functionalists claim that the same mind could be instantiated in wildly different hardware, neural or otherwise, as long as the same functional relations are present. These claims have sparked heated arguments about the adequacy of functionalism as a metatheory of mind, most of which seem to be aimed at convincing the reader either that functionalism is obviously right or obviously wrong. The considerations raised in the present article suggest that a better articulated critique is possible.

Using color experience as a paradigm example, I have argued that there is a metaphorical brick wall that limits the kind of knowledge about experience that can be derived from behavior. On one side of this epistemological wall lie objective facts about relations among experiences that are part of our shared cognitive culture and that can be studied behaviorally. On the other side lie subjective, private aspects of individual experiences that cannot. Theoretically, I have identified this limitation with the isomorphism constraint, claiming that behavioral science can specify the structure of experience up to the level of isomorphism, but no farther. This analysis is intimately related to functionalism insofar as it states that what can be known behaviorally are relations among experiences. Such relations are, by definition, within the domain of functionalism, so anything that can be known about experience from behavior (i.e., up to the isomorphism constraint) can be captured by a functionalist account. I have also argued, however, that the nature of individual experiences (beyond the isomorphism constraint) cannot be known from behavior, even in principle. These nonrelational aspects of experience lie, by definition, outside the domain of functionalism; they are underconstrained by relations among mental states. It seems that the failure of functionalism to provide an account of these aspects should be counted against its claim of fully specifying the nature of mind. It does not seem to be able to do the whole job.

It can do at least *part* of the job, however. Specifying conscious mental states to the level of isomorphism is nothing to be sneezed at. In the domain of color experience, it is sufficient to specify completely the shape and structure of color space, which includes everything we know about color experience, scientifically speaking. There is even a serious question about whether there is anything more that can be known within the methodological/epistemological limitations of science. I have argued, for example, that no theory of consciousness, even a causal biological one, will be able to produce a testable account of the nature of individual experiences because the behavioral aspects of the tests required would not support inferences beyond the isomorphism constraint. One might argue, therefore, that it is unfair to require functionalism (or any other theory of mind, for that matter) to be able to account for individual experiences, because such theories cannot be tested in any case.

The one way in which I have argued that we may be able to get some understanding of experiential qualities beyond the level of isomorphism is by the sort of "end run" I described previously. If within-subject experiments using biological interventions can enlarge the experiential equivalence class of color perceivers beyond individuals (e.g., just me) to groups of individuals (e.g., you and me), then I can infer that anyone within my equivalence class has the same color experiences as I do. Such equivalence classes lie beyond the isomorphism constraint because there could, in principle, be many of them within a population of perceivers who are behaviorally indistinguishable (see Fig. 7). Science could not objectively specify the nature of the underlying experiences within any of these equivalence classes, but I would have subjective knowledge for the experiences of members of my own equivalence class, just as you would for members of your own equivalence class. Thus it seems that functionalism may succeed in specifying experience as fully as objective science will allow, even though I, as a conscious experiencer of colors, may be able to get some further knowledge via attributing my own subjective experiences to my equivalence class as defined by combined evidence from the brain and behavioral sciences.

The indeterminacy of the nature of individual experiences within functionalism is a small shortcoming, however, compared to the fact that it fails to discriminate between my experiences and the complete lack of such experiences in an information-processing system that has the same causal relational structure among its processes, but no experiences of any kind. The color machine described above, for example, has the same representational color space as a normal human trichromat but, because of the implications of the color-room argument, it seems highly unlikely that such a machine would have color experiences of any sort. The causal isomorphism of its color representations to those of normal trichromats is sufficient to guarantee that it cannot be distinguished from a normal trichromat by behavioral means, but not that it has color experiences of any sort. This is a problem because it means that, in addition to having no account of the qualities of experience, *functionalism has no account of experience at all.*

There is an interesting parallel here between implementation and experience with respect to functionalism.[19] It has been recognized from the outset that functionalism treats mental phenomena as independent of their physical realizations: Any set of physical events will do, provided they have the right causal relational structure. One way of stating this situation is that functionalism induces a set of equivalence classes on objects in which all those with precisely the same causal relational structure among their internal states fall into the same equivalence class, despite potentially enormous differences in their physical implementations. A similar claim can be made about functionalism in the experiential domain. Functionalism appears to be independent of the experiential realization of mental phenomena in much the same sense: Any set of experiences – indeed, even nonexperiential computational states – will do provided they have the right causal relational structure. Functionalism thus establishes a set of equivalence classes on minds in which all those with the same causal relational structure among their experiences fall into the same equiv-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

alence class, despite potentially enormous differences in their experiential realization.

Notice a crucial difference between these two cases, however. Although functionalism asserts that minds are independent of their particular physical realization, it maintains that they must have a physical realization to meet the requirement of proper causal connections to the physical world. Functionalism is therefore merely indifferent about what particular implementation it has. However, in the corresponding case of experiential realization, functionalism does not demand that minds have one at all. There is no requirement corresponding to the causal connection to the environment that can be used to enforce some kind of phenomenal component as there is in the case of a physical component. *Experience thus does not have any intrinsic necessity within a functionalist framework, even though experience is the defining characteristic of mental life.*

The conclusion I draw is that functionalism is the appropriate, state-of-the-art theory of mind from the standpoint of purely behavioral science, but it falls short of providing a full account of mental events. This is perhaps not too surprising given that one of the cornerstones of functionalism is the irrelevance of physical implementation. Without physiology, what objective facts are there to constrain the nature of mind except behavioral ones? And, if behavior can reveal only relational aspects of mental events, then a theory of mind built solely on relations can be expected to do as well as can be done. Within-subject designs employing physiological interventions, natural or artificial, provide ways of going beyond these limitations, however. If such research is successful – and neuropsychological phenomena such as blindsight suggest that it will be – more comprehensive, physically based theories of mind will be required to account for the fact that certain changes in neural realization produce behaviorally detectable changes in experience within individuals.

The considerations in this target article cast serious doubt on the possibility of science being able to give a complete and testable explanation of the quality of color experience, or any other kind of experience for that matter. Behavioral science is limited by the isomorphism constraint in making between-subject comparisons: It cannot detect differences in the experiences of two individuals below the level of equivalent relational structure. Biological science is subject to the same limitation, because behavioral measurements are the only way in which it can compare experiences across individuals. The prospects for going further using biological techniques depend on (1) establishing the identity of underlying physiology to a level at which it becomes implausible to believe that experiential differences would arise from such minor physical differences, and/or (2) employing physiological interventions using within-subject designs. Even if both conditions can be met (and each will be difficult), there is no way to specify uniquely the qualities of particular experiences except by reference to one's own. Physiological identity can demonstrate only that two people's experiences should be the same, not what actual qualities they have in common. Within-subject designs can examine changes in experience, but cannot reveal to or from what they changed. Thus, explaining the qualities of individual experiences may be one mystery that will not yield its secret to the seemingly inexorable advance of science.

ACKNOWLEDGMENTS

The preparation of this target article was supported by grant 1-R01-MH46141 from the National Institute of Mental Health to the author. I wish to thank Alison Gopnik, Paul Kay, Eleanor Rosch, John Watson, John Searle, Bruce Mangan, Max Vellmans, Bernard Baars, Elisabeth Pacherie, C. Lawrence Hardin, Robert E. MacLaury, and two anonymous reviewers for their helpful comments on earlier drafts of this target article.

NOTES

Requests for reprints should be sent to Stephen E. Palmer at the Psychology Department, University of California, Berkeley, CA 94720-1650.

**1.** Although there is a great deal of consensus among color scientists about the primacy of these four chromatic colors, the matter is not completely settled. Indeed, Saunders and van Brakel (1997) have disputed various aspects of this idea in a previous target article in this journal, to which the interested reader is referred.

**2.** Many people object that green is not a primary because it is a mixture of blue and yellow. Although it is true that mixing blue and yellow paints or dyes generally produces some shade of green, people without experience mixing paints do not report that green looks like the composition of blue and yellow. In any case, virtually everyone given control over the degree of yellowness/blueness in a green test patch can easily find a setting where it looks neither yellowish nor bluish. This color is unique green.

**3.** The color space shown in Figure 2 is intended to be an abstract schematic model of human color experience rather than a particular one, such as Munsell space, OSA space, NCS space, or CIE space, each of which has somewhat different properties in terms of both the types of stimuli and the nature of the relations they represent. This model will be used to explicate a mode of argumentation with respect to empirical constraints on the inverted-spectrum argument rather than to demonstrate definitive results concerning fine points in terms of which alternative models of color differ.

**4.** Again, this generally accepted view of the dimensionality of color experience is not beyond reproach. Saunders and van Brakel (1997) have disputed that these are the only or the most appropriate dimensions of color experience, citing a number of dissenting voices in the research literature. The interested reader is refered to their article for further discussion.

**5.** It can be objected that there is no reason to suppose that such individuals are any different from normal trichromats in any biologically meaningful way. This would be true, however, only if there is no difference between L- and M-cones except the nature of the pigments they contain and the genetic codes that determine those pigments. If there are any other differences between L- and M-cones that are determined by other genes – for example, their frequency of occurrence, retinal distribution, or output relations to other cell types – then such individuals should not be biologically grouped with normal trichromats.

**6.** This presentation reflects what I take to be a plausible first-order story about color categories, one that may or may not be correct in detail. I present it to illustrate the logic of the claim that the structure of color categories breaks further symmetries of an empirically accurate color space, rather than as a definitive argument that they do so. Many aspects of the Berlin-Kay-McDaniel theory I discuss are open to question, as Saunders and van Brakel (1997) have argued at length.

**7.** The status of goluboi as a BCT is controversial, with some researchers endorsing its inclusion and others not (see MacLaury 1997a).

**8.** Tarski (1954) took a set theoretic approach to models by defining an *n*-ary relation among objects in terms of the set of ordered *n*-tuples of objects for which that relation holds. For instance, the binary *lighter-than* relation between two color experiences, *lighter-than* $(c_i, c_j)$, can be defined as the set of all ordered pairs of color experiences $\langle c_i, c_j \rangle$ such that $c_i$ is lighter than $c_j$. This



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

relation, in the sense of a set of ordered pairs, is then structurally preserved by the binary *higher-than* relation in a spatial model of colors if there is a function, $F$, that maps colors $(c_1, c_2, \ldots, c_n)$ to points $(p_1, p_2, \ldots, p_n)$, $F(c_i) = p_i$ and $F(c_j) = p_j$, such that $\langle p_i, p_j \rangle$ is a member of the *higher-than* set of ordered pairs of points if and only if $\langle c_i, c_j \rangle$ is a member of the *lighter-than* set of ordered pairs of colors. These conditions ensure that the structure of one set of objects and the relations among them are identical to the structure of another set of objects and the relations among them, without either the corresponding objects or relations being identical in any sense. (In the discussion of interobserver color transformations, the relevant mapping is from color experiences in one observer to color experiences in another observer.)

**9.** Formally, an isomorphism can be defined as a function, $F$, that maps a relational system, $S = \langle E, R_1, R_2, \ldots R_n \rangle$, onto another relational system, $S' = \langle E', R_1, R_2, \ldots R_n \rangle$, such that $\langle e_i', e_j' \rangle$ is an element of $R_k'$ iff $\langle e_i, e_j \rangle$ is an element of $R_k$ and $F(e_i) = e_i'$ and $F(e_j) = e_j'$. $E$ and $E'$ are simple sets of elements in the two relational systems (in this case, color experiences of two different observers), and $R_i$ and $R_i'$ are sets of $n$-tuples that constitute relations among the elements within the two relational systems (in this case, relations among color experiences of two different observers).

**10.** Transitivity requires that if $A$ is isomorphic to $B$, and $B$ is isomorphic to $C$, then $A$ is isomorphic to $C$. Symmetry requires that if $A$ is isomorphic to $B$ then $B$ is isomorphic to $A$.

**11.** The concept of isomorphism has a long and somewhat confusing history in psychology. Members of the Gestalt school discussed two different relations in terms of isomorphism (e.g., Koffka 1935; Köhler 1929). The clearer and less problematic use was to describe the "psychophysiological" relation between sensory experiences and underlying neurological processes in terms of having the same abstract structure. (We will have more to say about this relation later in this target article.) The other was to describe the relation between the physical world and underlying neurological representations. Here the authors introduced some confusion by implying that there was actual similarity between certain spatial qualities of a stimulus, such as a square, and the corresponding pattern of neurological activity, which might itself be similar to a square. A more realistic view of this psychophysical form of isomorphism was developed by Shepard and Chipman (1970) as a natural extension of Shepard's pioneering work in nonmetric multidimensional scaling (Shepard 1962a; 1962b). Shepard and Chipman called this relation "second-order isomorphism" to emphasize that, unlike some of the Gestalt ideas about isomorphism, the "first-order" properties of representations do not have to be the same or even similar. Thus the internal representation of a square does not itself have to be square or even remotely similar to a square, but, whatever it is, it must be more like the internal representation of a rectangle than that of a triangle or circle. (Strictly speaking, the "second-order" qualifier is unnecessary because mathematical isomorphisms are, by definition, abstracted from the "first-order" properties of their individual objects and concern only "second-order" relational structure.) A particularly clear and cogent early discussion of isomorphism in sensory psychology can be found in Hayek (1952, pp. 37–41). None of these uses of isomorphism is the same as the present one, however, which concerns the relation between the experiences of two different observers who may, in fact, have qualitatively different experiences but the same relational structure over those experiences.

**12.** By "causal" I do not mean a theory of consciousness that specifies the nonconscious neural events in the causal chain that ultimately leads to consciousness. In color perception, for example, it is clear that the registration of wavelength information by activity in the short-, medium-, and long-wavelength cones of the retina is not itself conscious but does lead to color experience somewhere later in the sequence of neural processing. I take it that the specification of this causal chain of neural events leading to consciousness is conceptually unproblematic, although the details are not yet understood.

**13.** I hasten to add that this analogy is quite imperfect. I particularly do not want to be taken as claiming that consciousness must be reduced to some causal physical mechanism – although it might – in the same way that biological processes can be reduced to physical processes involving DNA or that consciousness is like the "vital force" that will be eliminated from the scientific vocabulary once it is properly understood in mechanistic terms. Both of these conjectures are up for grabs. I offer the analogy only as a particularly clear example of the important difference between correlational and causal theories in a scientific domain.

**14.** There appears to be a fairly widespread belief that the neural correlates of experienced redness/greenness, blueness/yellowness, and blackness/whiteness are already known from single cell recordings of opponent cells in the retina and lateral geniculate nucleus of monkeys (e.g., De Valois & Jacobs 1968). Although these results are indeed suggestive, they do not support such an extreme conclusion; it is extremely unlikely that the neural correlate of color experience arises in precortical structures. The most that can be said is that there may be some later representation of experienced color that has a similar neural coding in terms of three opponent processes, as Hering (1878/1964) suggested from a phenomenological analysis.

**15.** We would not require that our entire brains be identical, for there may be many brain differences that would be irrelevant to our experiences of color, such as those reflecting different memories of autobiographical events, different motor skills, or different likes and dislikes. Our own experiences of color do not appear to vary over the course of our lives as these and many other factors change, although such conclusions rest on assumptions that can be challenged.

**16.** For example, the red-green dimension of color experience might not have been reversed by the drug but, rather, the set of associations to the relevant linguistic terms for reporting them. Grass would then still *look* green, but the observer would call it "red." Such differences could be ruled out in a number of ways. For one, observers could be shown normal and color-inverted pictures of objects with characteristic colors (green grass and red grass, red stop signs and green stop signs) and then be asked (1) which were normal and which were color inverted, and (2) the name of the color of the object shown. If the experiences were inverted, subjects would always be wrong in both tasks; if the linguistic labels were inverted, subjects would name the colors incorrectly but discriminate normal from reversed colors correctly. Another approach would be to determine which physiological structures were altered by the drug. If they were in the color pathways of the visual system, color experience would be the more plausible alternative; if they were in the linguistic or memory centers, a change in language would be more plausible.

**17.** A further methodological difficulty in this program of research is accepting the null hypothesis. That is, to eliminate a given neurological factor as relevant to color experience, one has to claim that it has no effect, and this is statistically problematic.

**18.** This is true for normal people under normal circumstances. However, in a variety of dissociative states, such as during hypnosis, fugue states, and episodes of multiple personalities, there appear to be subjectivity barriers within individuals. These are complex and interesting phenomena that will ultimately prove to be of great importance to understanding consciousness but are well beyond the scope of the present discussion.

**19.** I thank Alison Gopnik for pointing out this parallelism when we discussed an earlier draft of this target article.



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

# Open Peer Commentary

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

## How to compare color sensations in different brains

Werner Backhaus

*Theoretical and Experimental Biology, Department of Biology, WE 05, Freie Universität Berlin, 14195 Berlin, Germany.* **backhaus@zedat.fu-berlin.de**
**www.fu-berlin.de/backhaus**

**Abstract:** The qualitative and quantitative properties of color sensations and neuronal color coding are discussed in relation to physiological color exchanges and their evolutionary constraints. Based on the identity mind/matter thesis, additional physical measurements on color sensations are described that will allow us, at least in principle, to compare the qualitative properties of color sensations in different brains.

***1. Qualitative and quantitative properties of color sensations.*** Our color sensations consist of six elementary colors with different hue quality, brightness (lightness) being an additional common quality. These qualities can be compared and judged by introspection, at least within one brain (content-analytical experiments). The hues of the elementary colors can be named (red, green, blue, yellow, black, and white) and the intensity of their color brightness can be judged, as well as the amounts of elementary colors in any color sensation (see Backhaus et al. 1998; Backhaus 1996; 1998a; 1998e). Animals can learn, as a food signal, a specific elementary color that two or more colors have in common (through double or multiple color training) and they can choose the colors according to the respective amounts of this elementary color (bees: Backhaus 1995; also Backhaus 1994; 1997a; Backhaus & Kratzsch 1993). Color names are arbitrarily related to the elementary colors by learning and are hence unreliable for communicating of the qualitative (ontological) properties of colors in different brains.

On the other hand, the brightness ordering of the elementary colors appears to be fixed and thus provides more information for this purpose. Further (unconscious) properties are expected to exist in addition to the (conscious) properties determined by introspection; these can also be used to uniquely determine the elementary colors in different brains (Backhaus 1997b; 1997c; 1998a; 1998b; 1998e; 1999; in press; in prep.).

***2. Neuronal color coding.*** Neuronal color opponent coding (COC) appears to be the general strategy for transforming data from the photoreceptors into information concerning the elementary colors. COC neurons coding for red/grey/green, blue/grey/yellow, UV/grey/blue-green, blue/grey/UV-green, and so forth, as well as black and white have been shown to exist in humans and other animals. Inverted neurons, with swapped high- and low-frequency responses to monochromatic light exist, as well. This is probably to keep the excitation-dependent noise identical for all elementary colors (see Backhaus 1993; Backhaus et al. 1996). On the other hand, neurons that code exclusively for an elementary color have not been found in any species. Thus the quantity of elementary color must be encoded by the COC neurons (Backhaus 1998a; 1999). Figure 1 shows the COC model of color vision in the honeybee, which predicts and explains color similarity and color discrimination behavior in terms of the neuronal excitation values (Backhaus 1993; 1998c; 1998d). This model is in the process of being extended by temporal properties (Becker & Backhaus 1998; in press).

Nothing has so far been assumed about color sensations. Nevertheless, the results of the double-color training experiments

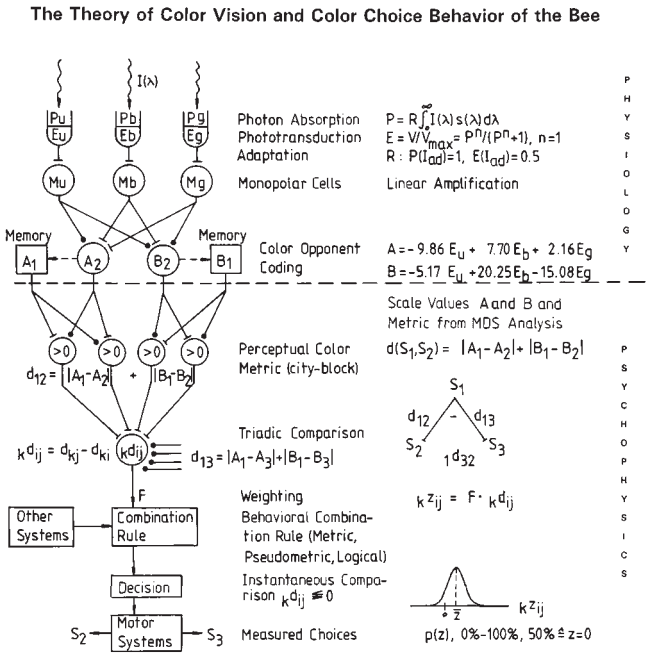



Figure 1 (Backhaus). COC model of neuronal color coding and color choice behavior for the bee. The upper part (above dashed line) is related to the physiology of the color vision system. The lower part is related to the psychophysics of color vision. I(λ): light intensity; u, b, g: photoreceptor spectral types; s(λ): spectral sensitivity; R: range sensitivity; P: absorbed photon flux; E: photoreceptor excitation; V: membrane potential; M: Monopolar cell; A, B: excitations of the COC neurons; $d_{kj}$: color difference between two color stimuli S; ${}_kd_{ij}$: judgment values; F: global scaling factor; ${}_kz_{ij}$: scaled judgment values; p(z): choice percentages (inverse z-transformed). The metric of subjective color space can be realized by simple neurons that receive inverted input from the COC neurons and corresponding color memories (from Backhaus 1993).

mentioned above could not be explained exclusively by COC neuronal activity; they appear to depend on a component (elementary color) of a further color representation. Whether this representation is conscious and colorful like ours can not be answered from behavioral experiments alone. These results nevertheless support the hypothesis that the judgments in ordinary color discrimination and color similarity experiments are performed unconsciously. Only color content analyses appear to rely on conscious judgments (Backhaus 1998e; 1999, see Fig. 2).

***3. Physiological color exchanges.*** From the biological point of view, we expect that the visual system of every species with color vision, including man, has adapted by coevolution to its environment for species-specific tasks (Backhaus 1992; Kevan & Backhaus 1998; Pielot et al., in press; Pielot & Backhaus, submitted). Thus color vision systems should be almost identical within a species, up to rare color vision deficiencies caused by mutations and the population spread related to genetical recombination. From this point of view, the problem raised in the target article reduces to two main questions: (1) Can there be physiological changes in color vision that have no expression in behavior and consequently will not have died out because of natural selection? (2) Is it possible to determine the properties of color sensations other than by introspection?

Regarding behaviorally silent color differences: Inverting the COC neurons would have the same result as swapping the corresponding elementary colors. Simultaneous inversion of the code and swapping of the elementary colors would leave the color vision system unchanged. Obviously, neither simple nor complicated transfer motions are possible if they disturb the ordering of color brightness. Only if all the elementary colors are swapped exactly


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

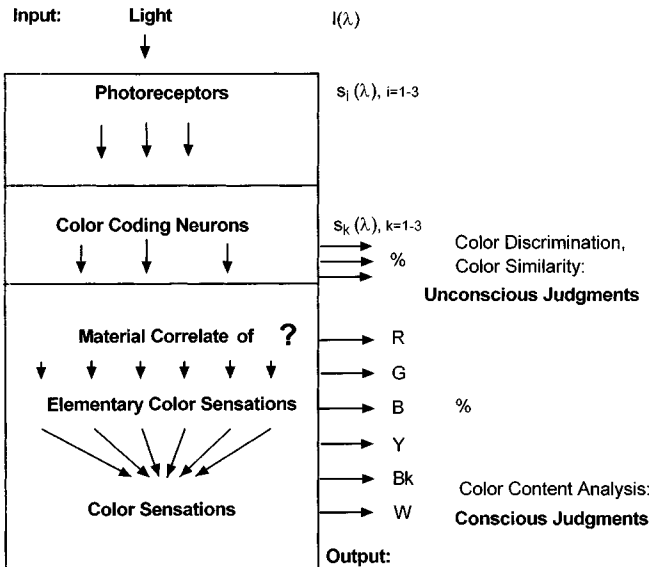



Figure 2 (Backhaus). Causal chain structure of color vision and supposed types of judgments (from Backhaus 1998e).

so that the color brightness scale is inverted (i.e., dark is reported as bright and *vice versa*), this change will remain undetected, at least in the experiments mentioned above. Nevertheless, from an evolutionary point of view this is rather unlikely, because several genes would have to mutate simultaneously. Once the genes for the color coding neurons and the elementary colors are known (as they already are for the photoreceptors), the unlikeliness can be calculated. Steps in between would die out by natural selection (see Pielot et al., in press; Pielot & Backhaus, submitted).

***4. Physical measurements will allow us to compare color sensations in different brains.*** Regarding introspection, consider physics, which is on the way to describing the world completely through a grand unified "theory of everything." This theory should also describe our conscious mental states, for example, the elementary colors and color sensations; otherwise, it will be incomplete. When the materials that are close to or even identical to the color sensations (see Fig. 3) are ultimately identified (the identity thesis), the problem of the nature of color sensations (experiences) in man and animals will reduce to the ordinary physiological problem of demonstrating that specific materials exist in respective parts of their brains.

Apart from conscious qualities, as directly (subjectively) observed by introspection, there should exist further unconscious physical properties (e.g., masses, quantum properties, etc.), which can be (objectively) measured (at least in principle). It should be possible to characterize color sensations physically per se, which should allow a deeper understanding of color sensations and will finally allow us to compare color sensations in different brains by objective methods (see Backhaus 1998b; 1999).

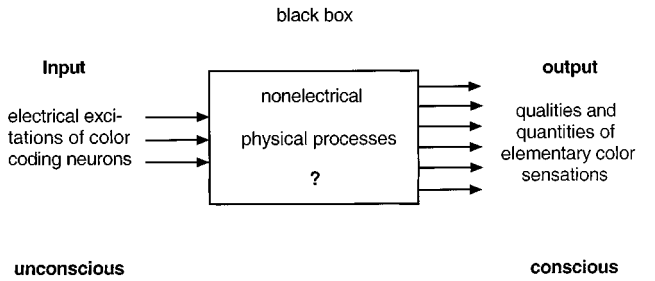



Figure 3 (Backhaus). Constraints for a physiological model of color sensations (from Backhaus 1998e).

## Color: How you see it, when you don't

Philip J. Benson

*University Laboratory of Physiology, Oxford OX1 3PT, United Kingdom.*
**philip.benson@physiol.ox.ac.uk** **www.physiol.ox.ac.uk/~pjb**

**Abstract:** It is worth considering whether particular behavioral measures from observers are ever consciously (or preattentively) transformed *a priori* so as to render inferences about them indistinguishable. This is unlikely, but recent experiments indicating color sensitivity and selectivity without visual awareness suggest that the distinction between what can and cannot be explained about color experience using behavioral responses may not be as obvious as Palmer concluded.

Consciousness is always accompanied by awareness. Awareness arises when we have overt access to information that can be used to influence behavior. Quantitative changes in behavior arising from associated phenomenal experience is often interpreted in terms of a functional neurological architecture. When a breakdown of the modularity is exposed in blindsight or apperceptive agnosia, for example, we can sometimes interpret quite profound residual skills using knowledge of alternative multiple parallel pathways. That being the case, when it is suggested that awareness need not be accompanied by consciousness we must remember that awareness is intimately influenced by the properties of the sensory stimulus. So consider the following.

When we asked an observer, known as G.Y., who has suffered complete degeneration of his left primary visual cortex, to judge the direction of a color stimulus moving in his blind field, performance accuracy was very good. We further found that he was sometimes conscious of a visual event, that "something happened," in his blind hemifield (Guo et al. 1998). But at no time did he see either the color of the object or its motion. His experiences are devoid of content, of qualia. Asked to name the hue – red, green, blue, or yellow – of a small patch of color as bright as its background, G.Y.'s hesitancy is understandable. However, his prompted verbal guessing was almost twice as good as he could have achieved by guessing alone (47%, $d' = 0.74$) – all very interesting for someone who has had no experience of chroma in his blind field in the 35 years since his accident. The pupillary light reflex of G.Y. and similar patients also shows sensitivity to stimulus-related color and motion (Stoerig et al. 1994), which provides an important indirect means of measuring the unconscious "perceptions" of his visual system.

Damage to a different brain area results in cortical color blindness and affects observers in a different way. Unlike G.Y., such patients can readily volunteer comments on their visual experiences of the same kind of stimuli, but they cannot overtly access their own representations of chromatic information that nevertheless allow them to process color contrast boundaries, for example. A completely achromatopsic observer, M.S., was unable to indicate verbally the position of the odd-one-out in a color test (Heywood et al. 1998). His saccadic eye movements to the odd color were also subject to overt guesswork, but were actually remarkably accurate.

The cognizances of observers like these have the potential to disclose important facets of color space and can be framed within Palmer's hypothesis. The question of color subjectivity, considered in terms of access-consciousness (Block 1995), is not, I believe, a *prima facie* reason to disqualify between-subjects designs. Palmer's suspicions have to be testable. Are they? I do not think we have yet journeyed into an experimental or philosophical cul-de-sac, but we have to assume that the fundamental Lockean question is worth asking. There are several methods at our disposal – direct, indirect, *and* behavioral – that allow us to speculate on the essence of color similarity judgments and color categories that might be made without awareness of hue, albeit at a fairly crude level. A damaged or chronically underused color space may become distorted, leading to tell-tale errors when it is accessed. The time of injury, and the frequency of participation in intense experimenta-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

tion might be relevant because there may be a strong element of acquiring the skills to associate subthreshold experiences with properties of color, motion, and form. Paradoxical neuropsychological evidence is there to furnish us with an operationalized interpretation of our normal behavioral repertoires and experiences.

ACKNOWLEDGMENTS

G.Y.'s color naming experiments were carried out in collaboration with Kun Guo. Preparation of this commentary was supported by the United Kingdom Medical Research Council.

## Jack and Jill have shifted spectra

Ned Block

*Department of Philosophy, New York University, New York, NY 10003.*
**nb21@is5.nyu.edu www.nyu.edu/gsas/dept/philo/faculty/block**

**Abstract:** There is reason to believe that people of different gender, race or age differ in spectra that are shifted relative to one another. Shifted spectra are not as dramatic as inverted spectra, but they can be used to make some of the same philosophical points.

***How important are the facts that may block an actual inverted spectrum?*** The inverted and absent qualia arguments against functionalism that Palmer invokes are well known in the philosophical literature,[1] although the wealth of empirical considerations he so convincingly describes are a new and welcome contribution. However, do the empirical issues really matter for the challenge that the possibility of an inverted spectrum poses for functionalism? My answer is: only in a subtle and indirect way. For if there *could be* creatures who have color experience or at least visual experience for whom the relevant empirical points do not apply, then functionalism is refuted even if *human* spectra cannot be inverted. For example, perhaps it is a consequence of the human genome that humans tend to be unwilling to give a name to light green, even though they are willing to give a name to light red ("pink"). Still, there *could* be people whose visual experience is similar even if not identical to ours who find it equally natural to give names to light green as to light red. (Perhaps we could even genetically engineer humans to become so willing.) If this and other relevant asymmetries could fail to obtain in people (or other creatures) who have color experience (or at least visual experience), then the mere *possibility* of an inverted spectrum shows that functionalism is false (Block 1990; Shoemaker 1982). It would be a strange claim that the phenomenal character of *our* visual states has a functional/computational essence, even though there could be *other creatures* who have visual experiences that have phenomenal characters that lack that essence.

However, just because it *seems* that there *could be* living beings, even if not humans, who could have an inverted spectrum, does that show that such creatures really could exist? This, it seems to me, is where the empirical considerations come in; they move the burden of proof to the functionalist. Three transformations of color space survive the asymmetries that are known to be determined to the visual system itself. As Palmer emphasizes, the further asymmetries that may block an indetectible inverted spectrum may well be determined by a later cognitive stage of processing. Currently available empirical considerations, therefore, do not count against the possibility that there *already exist* cases of inversion of color experience. Suppose, however, that the color naming and categorization asymmetries do in fact reflect asymmetries in color experience. Should it not be possible for there to be people (or creatures of some sort) for whom light green looks as salient (and therefore nameable) as light red? The difference between such creatures and humans seems trivial, one of the sort that might for all I know obtain between my neighbor and me. This triviality puts the burden on the functionalist.

Philosophers make much of the distinction between a functional state or process and its realizer or implementer (Block 1980). For example, a computational process for finding square roots might be realized or implemented by a mechanical (gears and wheels) system, or by an electronic or biological system (a human doing the computing). In the battle between Computation and Biology that has defined the last 40 years of controversy in cognitive science, the Computationalists say that qualia (i.e., the phenomenal feels of sensory experience) are functional states, whereas the Biologicists say that qualia are the biological realizers or implementers of those states. The mere possibility of an inverted spectrum, if it can be established, gives the Biologicists an advantage over the Computationalists when it comes to qualia.[2]

***Shifted spectra.*** Functionalism is only one of the targets of arguments from inverted and absent qualia. Another is representationism. Representationism is the view that the phenomenal character of an experience is (or at least is determined by) its representational content. That is, the phenomenal character of an experience as of red consists simply in the experience representing something *as red.* (This view is taken by Dretske 1995; Harman 1996; Lycan 1996; Tye 1995, and, less clearly in McDowell 1994b.) These representational contents are usually supposed to be nonconceptual, as distinct from the contents of thoughts. (Nonconceptual contents of perception are shared by people and animals, even if the people can reason with these contents, whereas the animals cannot. I will not try to spell out the concept of nonconceptual content further; see Crane 1992.) If you and I can have inverted phenomenal characters even though our experiences represent the same color, representationism is refuted.

But although the kind of inverted spectrum needed to refute functionalism requires behavioral (and functional) isomorphism (as Palmer notes), representationism can perhaps be refuted empirically without these isomorphisms, as I shall now argue. One interesting empirical argument appeals to the fact that color vision varies from one *normal* person to another. Two normal people chosen at random will differ half the time in peak cone sensitivity by 1–2 nm or more. (More precisely, the standard deviation is 1–2 nm; see Lutze et al. 1990.) This is a considerable difference, given that the red and green cones differ in peak sensitivities only by about 25 nm. Moreover, there are a number of specific genetic divisions in peak sensitivities in the population that are analogous to differences in blood types (in that they are genetic polymorphisms, discontinuous genetic differences coding for different types of normal individuals). The most dramatic of these is a 62% – 38% split in the population of two types of long-wave (red) cones that differ by 5–7 nm (roughly 24% of the difference between the peak sensitivities of red and green cones; Neitz et al. 1993). This characteristic is sex-linked. The distribution just mentioned is for men. Women have smaller numbers in the two extreme categories and a much larger number in between. As a result, the match on the Rayleigh test (described below) "most frequently made by female subjects occurs where no male matches" (Neitz & Jacobs 1986).

These differences in peak sensitivities do not show up in normal activities, but they do reveal themselves in subtle experimental situations. One such experimental paradigm uses the anomaloscope (devised in the nineteenth century by Lord Rayleigh), in which subjects are asked to make two halves of a screen match in color, where one half is lit by a mixture of red and green light and the other half is lit by yellow light. The subjects can control the intensities of the red and green lights. Neitz et al. 1993 note that "people who differ in middle wavelength sensitivity (M) or long wavelength sensitivity (L) cone pigments disagree in the proportion of the mixture primaries required" (p. 117). That is, whereas one subject may see the two sides as the same in color, another may see them as different, for example, one redder than the other. Furthermore, variation in peak sensitivities of cones is just one kind of color vision variation. The shape of the sensitivity curves likewise varies. These differences are the result of differences in ocular pigmentation, which vary with "both age and degree of skin pigmentation" (Neitz & Jacobs 1986). There is also considerable


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

variation in the amount of light absorption by preretinal structures, and this factor also varies with age.

So all should agree that a red object probably will not look exactly the same, color-wise, to me as to you, especially if we differ in age, gender or race. (I emphasize gender, race, and age to stifle the reaction that one group should be regarded as normal and the others as defective.) We now have the beginnings of an argument against representationism. Jack and Jill have experience that represents red things as red and scarlet things as scarlet even though they experience both red and scarlet slightly differently.

This argument does not quite work, however, for representationists could reply that the representational contents of Jack's and Jill's color categories may differ, too, so there is still no proven gap between representational content and phenomenal character. "Color categories?" you say. "I thought the representationist was talking about *nonconceptual* contents." True, but the representationist has to allow that our visual experiences represent a scarlet thing as red as well as scarlet. For we experience scarlet as both red and as scarlet. We experience two red things of different shades as having the same color, though not the same shade, so a representationist has to concede a component of the representational contents of experience that capture that fact about experience. The representationist must allow representational content of both color and shade. Furthermore, pigeons can be conditioned to peck at things of a certain color, as well as of a certain narrow shade. Even if the pigeon lacks color concepts, it has something short of them that involves groups of shades as well as shades, red as well as scarlet. Let us use the term "category" for this aspect of the nonconceptual contents that is conceptlike but can be had by animals that cannot reason with the contents. Now we can see why the argument I gave does not quite work against the representationist. Jack's and Jill's experiences of a single red fire hydrant may differ in phenomenal character *and* also in representational content, because, say, Jack's visual category of red may include a shade that is included instead in Jill's visual category of orange. So we do not yet have the wedge between phenomenal character and representational content.

The way to get the wedge is to apply the argument just given to shades rather than colors. Let us co-opt the word "aquamarine" to denote a shade of blue that is as narrow as a shade can be, one that has no subshades. If Jack's and Jill's visual systems differ slightly in the ways that I described earlier, then perhaps aquamarine does not look to Jack the way it looks to Jill. Perhaps aquamarine looks to Jack the way turquoise (a different minimal shade, let us say) looks to Jill. But why should we think there is any difference between the representational contents of Jack's experience of aquamarine and Jill's? They both acquired their categories of aquamarine by being shown (let us suppose) a standard aquamarine chip. It is that objective color that their (different) experiences of aquamarine both represent. The upshot is that there is an empirically based argument (that one might call "shifted spectra") for a conclusion that, though not as dramatic as an inverted spectrum, has much the same consequences for representationism and for the question of whether there are uniform phenomenal characters corresponding to colors. There may be small phenomenal differences among normal people that do not track the colors that are represented.[3] Genders, races, and ages may differ by shifted spectra.

I mentioned above that there is an objection to my first point against representationism: Jack's visual category that represents red may include a shade that is included in Jill's visual category that represents orange. The present point is that the same argument does not apply to minimal shades themselves.

This possibility should not disturb the functionalist as much as the representationist, however, for even if there are phenomenal differences among representationally identical experiences as just supposed, the phenomenal differences might be revealed in subtle empirical tests of the sort I mentioned. That is, perhaps shifted spectra always result in different matches on a Rayleigh Anomaloscope or other devices. But shifted spectra would still count against representationism.[4]

Of course, the argument about the irrelevance of the actual facts given in the first part of this commentary also applies to the second part. It is certainly *conceptually* possible that two normal people looking at a single color patch have experiences that both represent it as aquamarine, yet the two have slightly different experiences of it. And the conceptual point in this case is clearer than in the inverted spectrum case, because there is no issue of whether someone with a different structure of color space has color experience at all. The upshot is that the representational contents of color experiences could be the same even though the phenomenal characters are slightly different; representationism is therefore false.

ACKNOWLEDGMENT

I am indebted to Paul Boghossian, Tyler Burge, and Susan Carey for comments on an earlier draft.

NOTES

**1.** The qualia-based arguments against functionalism that Palmer elucidates appeared first, I believe, in Block and Fodor (1972), who coin the terms "inverted qualia" and "absent qualia." The point that functionalism can characterize mental states up to the level of isomorphism was made by Shoemaker (1975). Shoemaker (1982) also emphasizes the advantages of the "within subjects" version of the inverted spectrum hypothesis. A spectrum inversion example involving two eyes is given by Byrne and Hilbert (1997). A version of the absent qualia argument that Palmer bases on Searle's (1980) Chinese Room argument was given in Block (1978), the main difference being that instead of the functional relations being realized in a single person, Block (1978) uses a group of people communicating electronically.

**2.** See Rey (1997, p. 311–12) for a functionalist response.

**3.** Some may wish to avoid this conclusion by insisting that colors are not real properties of things, that our experience ascribes phenomenal properties to physical objects that the objects do not and could not have (Boghossian & Velleman 1989; 1991). But this cannot be said by representationists, because it would bring back unreduced phenomenal properties – which it is the essence of their position to reject. (See Shoemaker, 1994, for arguments that objects do have phenomenal properties.)

**4.** There is a complication that I cannot treat fully here. If you regard a certain mixture of red and green as matching the aquamarine chip, but I do not, then our categories of aquamarine are applied by us to different things, and in that sense have different extensions. I do not regard this as showing that our categories have different representational contents, because representational contents have to do with what objective colors are represented and the example given exploits an indeterminacy in objective color. There is no determinate answer as to whether the color of the mixture of red and green (that matches aquamarine according to me but not you) actually *is* aquamarine. See Block (1999) for a great deal more discussion of this issue.

## Scaling the metaphorical brick wall

Michael Bradie

*Department of Philosophy, Bowling Green State University, Bowling Green, OH 43403.* **mbradie@bgnet.bgsu.edu**

**Abstract:** Palmer argues that functionalist accounts of the mind are radically incomplete in virtue of a "metaphorical brick wall" that precludes a complete treatment of qualia. I argue that functionalists should remain unmoved by this line of argument to the effect that their accounts fail to do justice to some "intrinsic" features of experience.

Intersubjective comparison of qualia run afoul of what Palmer calls the "metaphorical brick wall" that separates the subjective aspects of our experiences (private to each of us) from the objective aspects of our experiences, which are publicly accessible to others (sect. 4, para. 2). Functionalist accounts of mental experiences are capable of dealing with the objective aspects but cannot get a grip on the subjective aspects. But, these subjective aspects, so Palmer argues, are central to our concept of what having mental experi-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

ences is all about – they are "intrinsic" to the experience (sect. 2.1, para. 2). So, he concludes, functionalism is, at best, an incomplete account of our mental experiences, in particular, of our color experiences. (sect. 4, para. 8) Functionalism appears to be unable to distinguish between cases where I see red and you see blue (the "inverted spectrum" or "inverted color" problem). Indeed, functionalism seems unable to distinguish between human beings who experience red from robots, whose functional responses are indistinguishable from ours but who have no "experiences" at all. In short, functionalism leaves the "redness" out of our seeing red.

Palmer concludes: "*Experience thus does not have any intrinsic necessity within a functionalist framework, even though experience is the defining characteristic of mental life*" (sect. 4, para. 7). If we understand "experience" here to mean what philosophers of mind typically refer to as "qualia," then the complaint against functionalism is that it cannot account for qualia. But, so the argument goes, what we call our mental life is irreducibly qualia-laden. So functionalism must be radically incomplete.

Functionalists should remain unmoved by such considerations on the grounds that the alleged qualia that are supposedly part of the intrinsic nature of our mental lives remain hopelessly inaccessible. It thus behooves the critic of functionalism to provide reasons for thinking that there are any such features of our mental life that are left out of sophisticated functionalist accounts.

Palmer agrees that interpersonal comparisons of qualia are out of the question because of the "metaphorical brick wall" that isolates each of our subjectivities from the scrutiny of others. He suggests, however, that we can, in principle at least, do an "end-run" around this obstacle by appealing to intrapersonal comparisons of our experiences at one time with our experiences at another. The claim is that there is no subjectivity barrier within an individual. (sect. 3.6, para. 5) This is problematic. Someone with a heightened sense of "qualia" recollection (such as the character in Marcel Proust's *Remembrance of things past* whose reverie is set off by the smell of madeleines) might claim to be able to recall qualia from past experiences, compare them with qualia of present experiences, and thereby be able to determine whether like *objective* circumstances *were* or *were not* producing like qualia (Proust 1934). On what basis would such a claim rest? One can only claim that these qualia are the same or different on the basis of some external or publicly accessible criteria. I take it that this is one of the upshots of Wittgenstein's "private language" argument in the *Investigations* (Wittgenstein 1972, para. 258 and 265). So, for the purpose of "qualia- comparison," my prior self is as alien a being to my present self as any other mind. If this is right, some other device must be invoked to persuade us that there is something "intrinsic" missing from the functionalist account of color experiences.

## A wiring demon meets socialized humans and calibrated photometers

Michael H. Brill

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

**Abstract:** Ignoring consciousness, I apply Palmer's ideas to a photometer, for which calibration is analogous to socialization of humans to agree in color. Some attributes of the photometer – such as its aperture – do not need to be known because their values are transparent to calibration. But a writing demon can wreak havoc if it permutes measured values before interpolation completes calibration – as happens in Palmer's color rewirings.

Meet the wiring demon (WD), who lives to mix up the "wires" in any functional system, subject to rules provided by us. Palmer examines some possible works of WD, in the complicated and irregular system called color vision. To simplify and to add concreteness, I will first discard consciousness from the discussion and then retreat from color vision into the one-dimensional world of the laboratory photometer. Then, I will allow WD certain liberties and try to correct them by recalibrating my photometer. Perhaps the reader will agree that my contest with WD runs the gamut of Palmer's substantive issues without getting into unanswerable questions that have caused headaches since Plato wrote the *Timaeus*.

***1. More on isomorphisms, less on consciousness.*** In agreement with many texts, *Merriam-Webster's Collegiate Dictionary* (10th ed., 1995) defines "isomorphism" as "a one-to-one correspondence between two mathematical sets." This is a much looser definition than Palmer's (attributed to Tarski), which requires a prespecified "structure of relations" to be preserved by the mapping. What kind of structural relations are relevant? If the relation is distance, the mapping is called an isometry. If each small-enough neighborhood is mapped to another neighborhood, the mapping is called a diffeomorphism. Without a prespecified relation, the dictionary definition of isomorphism gives the widest latitude to WD, who can interchange the final (private) percepts of all the colors. Subject to certain restrictions, such as not mapping small measurement errors into large color errors, any such isomorphism is behaviorally undetectable. The human, having been socialized to attribute "red" as others do, is assumed to *invert* the isomorphism. Note that consciousness is not part of the picture.

Suppose WD attacks some stage prior to the final, ineffable, "physical-states-to-percepts" transformation. Then (again irrespective of consciousness) Tarski's relational provisos on "isomorphism" become significant. For example, if WD reverses basic RED and GREEN prior to the presumed derivation of PURPLE from RED AND BLUE, then (and only then) one gets Palmer's PURPLE = GREEN AND BLUE. In that case, the only isomorphism left for WD would have to preserve all the fuzzy-logic-linked pairs of the basic colors.

***2. Hidden life of a photometer.*** Calibrating a photometer inverts an isomorphism (between photometric scales) in much the same way that human socialization could compensate for a WD in color vision. In fact, the units of luminance betray certain internal parameters of the photometer that are ignored in measurements because they are "calibrated out." Luminance (cd/m^2) is the number of lumens ("visible watts") per steradian per square meter (see VESA 1998). The square meter refers to a diffusely emitting surface; the steradian refers to the solid angle, from a point on the surface, subtended by the aperture of the photometer. For a uniformly emitting plane, the distance of the photometer does not matter: The photometer intercepts less light from each point on a more distant plane, but the number of "seen" points is greater in exact compensation. But the aperture should matter: A smaller aperture means the photometer gets less light. Why is it not necessary to know the aperture to make a luminance measurement? The answer is because the luminance scale of the photometer has been calibrated using one or more known-luminance surfaces. Each photometer has a private world of received light, but it is calibrated to agree with other photometers.

When the photometer aperture is calibrated out, the isomorphism being inverted is a scaling from $R^+$ to $R^+$, between photometers, and the relation it preserves is the ratio between two readings. Other isomorphisms, too, can be calibrated out. For a smooth, strictly monotonic nonlinearity arising from the light detector, the preserved relation is the ordering of outputs, greatest to least. Also, neighborhoods are mapped to neighborhoods – a good thing, because small, random measurement errors map to small output errors. In contrast, a WD allowed free rein in its input-output permutation would produce large random output errors that would be uncorrectable by recalibration. Similarly, if WD preserved order but not smoothness in the rewiring, interpolation of readings would be impossible and it would therefore be impossible to recalibrate the device in a finite time.

When WD (alias the difference between photometers) is benign, it is possible to calibrate using a few standard light sources.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

These induce a number of fiducial readings. (Dare I say "basic luminance categories"?) Secondary (derived) readings could then be interpolated from pairs of the fiducial readings. If my interpolations were inaccurate, my photometer's readings might not agree with others. The agreement would be disastrous if I calibrated the fiducials of photometer 2, used those of photometer 1 to derive secondaries, and adopted these secondaries directly for photometer 2. That would be analogous to Palmer's PURPLE = GREEN AND BLUE. To see the analogy more clearly, imagine that WD permutes only the primary measurements, but that the calibration maps the whole luminance scale back to photometer 1, including the interpolated values. To conclude, I am not sure how this partial-calibration exercise could enlighten either a user of photometers or (by extension) a color scientist.

***3. Outlook.*** I have tried here to reinterpret the facts in Palmer's discussion so as to avoid unanswerable questions in favor of practical answers. I have not entirely succeeded, for metrology itself has an intrinsic philosophical question. Calibration is, in theory, a regress of photometers and light sources. Where does it begin?

## Subjectivity is no barrier

Alex Byrne

*Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139.* **abyrne@mit.edu**
**web.mit.edu/philos/www/byrne.html**

**Abstract:** Palmer's "subjectivity barrier" seems to be erected on a popular but highly suspect conception of visual experience, and his "color room" argument is invalid.

Palmer beautifully articulates a view that many philosophers and psychologists have found compelling: In attempting to understand the mind, scientists face an impenetrable "subjectivity barrier," behind which lies the "nature of [our] experiences themselves" (sect. 2.1, para. 2 and 3).[1] And although Palmer sometimes sweetens the pill with such locutions as "scientists would never know with certainty" (sect. 2.1, para. 1), and "it may not be possible to be sure" (sect. 2.2, para. 3), it is perfectly clear that the conclusions of his arguments are – setting futuristic brainometers aside – that *we* (not just "behavioral scientists") have *absolutely no reason to believe* (not just no "objective knowledge") that others have experiences with the same "intrinsic qualities" as our own, or even that they have any experiences at all (sect. 2.5, para. 8). Thus, what might seem at first glance a sober essay curbing the pretensions of science to explain consciousness is in fact a radically skeptical manifesto.

According to Palmer, we can at best know about the relational structure of others' experiences – the similarities and differences they bear to each other. (For brevity's sake I shall ignore his further claim that we cannot know that others have experiences.) Skepticism follows because Palmer holds that fixing this relational structure does not fix the "sensory qualities" of experiences.

Palmer's main examples of a difference in sensory quality with no difference in relational structure are variants of the usual "spectrum inversion" case (Shoemaker 1981). No doubt some commentators will attack Palmer's argument by questioning whether our color space is as symmetric as Palmer makes it out to be. But this response does not dig deep enough, because on Palmer's conception there may well be sensory qualities "alien" to us that could occupy the relational structure of our experiences.

A potentially more promising response disputes Palmer's crucial claim that sensory qualities can vary independently of relational structure. This response subdivides into two: the first agrees with Palmer that interpersonal comparisons of sensory qualities make sense (Clark 1993; Hilbert & Kalderon, in press), whereas the second denies that they do (Stalnaker, in press).[2] As I am unconvinced that the response just mentioned can be made to work, I shall try a different tack, and present a natural conception of visual experience on which it is hard to get Palmer's skeptical worries going.[3]

Imagine seeing a red object – the proverbial ripe tomato. Focus your attention on the salient property of the tomato that it shares with cherries and strawberries. That property is redness, right? And of course red objects – like tomatoes and cherries – look more similar, with respect to color, to orange objects than they do to green objects. These similarity relations among *colored objects* induce corresponding similarity relations among *our experiences of color*, rather than the other way around. It is not true that red objects are more similar to orange objects than they are to green objects *because* our experiences of red objects are more similar to our experiences of orange objects than they are to our experiences of green objects. Rather, the explanation goes the other way: Our experiences are similar because the *objects* of the experiences are similar. This is because when people have visual experiences they are only aware of *what visually appears to them*, or of *what their experience represents*: as it might be, the presence of a red tomato. They are not aware (at least, not without further effort) *of* their experiences. And similarly with the "nature of the experiences themselves": The perceived nature of the *objects* of perception (e.g., the redness of the tomato) explains the nature of *perceptions* of objects. If a person has a visual experience that represents the tomato as being *red*, then nothing else needs to be added to this experience for it to have the distinctive "sensory quality of redness" (sect. 2.1, para. 3) that Palmer thinks is hidden from scientific enquiry. Therefore, because there appears to be no *special* problem about knowing whether objects visually appear red to people, there is no special problem about knowing the nature of others' visual experiences, and thus there is no "subjectivity barrier."

Palmer may have been seduced by the following perennially appealing argument for thinking that something else needs to be added to a visual experience that represents something as red for it to have the "sensory quality of redness." Imagine having a "red" afterimage. Perhaps unreflectively you would be inclined to call the distinctive property of the image "red," but surely the image cannot really *be* red – red is a property of physical objects like tomatoes, not "mental objects" like images. So call the property of the image "R" instead. But obviously R is present when you see objects, like ripe tomatoes. And because "mental objects" like images can have R, it does not seem likely that R can ever be a property of a physical object like a tomato. So when you see that a tomato is red, you are aware that the tomato is red *and* that some image-like thing has R. It is the presence of an R-image that gives your experiences of red objects (and certain afterimages) their distinctive "sensory quality"; similarly, the distinctive sensory quality of your experiences of green objects is due to the presence of a G-image. And now, of course, the question naturally arises whether others' experiences of red objects are in fact attended by G-images rather than R-images.

As tempting as this reasoning is, it is wrong. If R is to explain the sensory quality of your experience of a ripe tomato, then it is not sufficient that the experience involve an image that *has* R: It must *visually appear to you* that the image has R (imagine that even though the image has R, *it appears to have G*; in that case you would have an experience with the sensory quality distinctive of your experiences of green objects). But now the alleged fact that the image *has* R is doing no explanatory work: The sensory quality of your experience is solely explained by the fact that *it appears to you that the image has R*, irrespective of whether the image has R. So the introduction of R was an idle wheel – *redness* would do the job just as well. Your afterimage experience was a kind of hallucination: It visually appeared to you that something was *red* (that is what gave your experience its distinctive sensory quality), but *nothing* in the scene before your eyes was red (that is why it was a kind of hallucination). Furthermore, although it seemed to you that there was an image floating before your eyes, in fact there was no object – not even a mental one – there at all.

Finally, as has been pointed out numerous times (e.g., Cope-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

land 1993), the Chinese room argument is fallacious. The conclusion concerns the *system* (it cannot understand Chinese), but the premise concerns a *part* of the system (the man does not understand Chinese). The argument is an instance of "*x* is not *F*, *x* is part of *y*, therefore *y* is not *F*," and so is invalid.[4] Thus Palmer's "color room" argument fails to show anything whatever about functionalism and experience.

NOTES

**1.** Palmer cites Wittgenstein as a supporter, but I think the reverse is true: The view Palmer holds is one that Wittgenstein argued against. Frege – the inventor of modern logic and one of the founders of analytical philosophy – is a much better candidate (see especially 1918/1988, and also 1884/1950, sect. 26, where Frege uses the dual system analogy of the target article [sect. 2.3, para. 11 and 12]).

**2.** Stalnaker draws a helpful analogy with Von Neumann-Morgenstern utility theory, which assigns utility scales to people who can only be compared intrapersonally, not interpersonally.

**3.** See Armstrong 1968; Byrne and Hilbert 1997; Dretske 1995; Harman 1990; Lycan 1996; Tye 1995. There are important dissenters, in particular, Block 1990 and 1995.

**4.** Searle's response is to let the man perform all the symbolic manipulations in his head, but this appeals to the inference pattern "*x* is not *F*, *y* is part of *x*, therefore *y* is not *F*," which is also invalid.

## Why asymmetries in color space cannot save functionalism

Jonathan Cohen

*Department of Philosophy, Rutgers University, New Brunswick, NJ 08901.*
**joncohen@ruccs.rutgers.edu**
**ruccs.rutgers.edu/~joncohen/cohen.html**

**Abstract:** Palmer's strategy of saving functionalism by constraining spectrum inversions cannot succeed because (1) there remain many nontrivial transformations not ruled out by Palmer's constraints, and (2) the constraints involved are due to the contingent makeup of our visual systems, and are therefore not available for use by functionalists.

The possibility of spectrum inversions has been taken to threaten the viability of individuating conscious states functionally – by their connections to perceptual input, behavioral output, and other mental states; such cases hypothesize token states that are functionally identical but nonetheless differ in phenomenal character. Palmer attempts to respond to this threat by pointing to substantive constraints that he says putative inversions would have to meet. However, he ultimately gives up on functionalism for two reasons. First, he is skeptical that his constraints rule out all candidate inversions. Second, he is convinced by absent-qualia objections alleging that no amount of functional connection is sufficient for conscious experience (sect. 2.5).

I share Palmer's pessimism about the prospects for an adequate functionalist account of color experience. Indeed, I think things are even worse for the functionalist than Palmer allows. For even if Palmer's constraints worked out better than he supposed, this would still not be enough to save functionalism from the inverted spectrum.[1]

The strategy Palmer considers involves pointing to functional properties of phenomenal characters of color experiences, thereby bringing phenomenal characters under the explanatory scope of functionalism. Palmer argues that phenomenal characters stand with each other in certain empirically salient relations (e.g., of similarity and composition), which must be left invariant by any spectrum inversion (by definition, inversions must leave functional connections fixed), and that this restriction radically constrains the space of putative inversions. Indeed, he suggests that if the relations considered in section 1.4 reflect properties of the phenomenal characters themselves (not sociolinguistically transmitted conventions), there can be no nontrivial isomorphisms, so functionalism would be saved.

Unfortunately, Palmer's restrictions on spectrum inversions are both too weak and too strong to save functionalism. They are too weak because they do not preclude all nontrivial transformations (even granting the applicability of the constraints of sect. 1.4). For, although Palmer does not discuss them, there remain transformations that map green to green, red to red, and so forth, but effect slight reorganizations *within* each color category. For example, there are transformations that move each unique color onto a close neighbor within the same color category, while leaving the ordering and metrical relations between colors fixed (slight rotations, or local stretchings/squeezings, of the color space). Could these transformations be ruled out by adducing further constraints? They'd better not, for some such transformation seems to be actual: As discussed by Hurvich (1981, pp. 222–23), there is a non-trivial interpersonal distribution of the loci of unique hues. Thus, even if all of Palmer's constraints are met, interpersonal differences are still possible (and, in some cases, actual) between the phenomenal characters of experiences of the same stimulus under the same viewing conditions.

At the same time, Palmer's constraints are too strong to save functionalism, because the relational properties involved are due to contingent properties of our visual systems, and hence are not within the reach of functionalism. For, as Palmer points out (sect. 4, para. 6), it is because functionalism individuates states in terms of functional roles irrespective of these roles' neuroanatomical implementation that functionalists can (and type-identity theorists cannot) say that a given state is shared by actual and possible creatures with widely divergent brain structures. But the relations Palmer wants to leave fixed lack this generality: The generally accepted accounts derive these similarity and compositional relations from the contingent makeup of the minds/visual systems of normal trichromats.[2] Even if this makeup is shared by macaques, it is certainly not shared by all the actual and imaginable creatures (e.g., dolphins, super-sophisticated Martians) to whom we might want to attribute color experiences. Therefore, functionalists can only advert to Palmer's relational constraints if they are prepared to become species-chauvinists. Presumably they are not, though, or they would not be functionalists.

The modal reflex of this problem is quite serious: If it is to specify the essence of color experiences – to list the features states must have to count as color experiences, rather than just the features color experiences happen to have – functionalism must provide an analysis of the color experiences of any metaphysically possible creature who has them. For this reason, functionalism is vulnerable to spectrum inversions not just between actual human beings, but between any two metaphysically possible creatures. And, as noted, Palmer's constraints lack this modal power. Although our color experiences might be uniquely constrained by the asymmetries Palmer considers, (1) there seem to be metaphysically possible subjects of whose color experiences this is not true, and (2) it is apparently contingent that we are not such creatures.

If this is right, functionalism can only state metaphysically contingent truths about color experience. Contingent truths can be interesting, but they do not serve the scientific goal of laying bare essences.[3]

NOTES

**1.** I have two reasons for pressing these points against Palmer even though he ultimately gives up functionalism. First, I think he underestimates the force of inverted spectrum worries. Second, I intend my criticism of the strategy to speak to others (e.g., Tye 1995, p. 202–205) who apply analogous strategies more confidently, and claim to have answered the threat of inverted spectra.

**2.** See Hurvich (1981) for an explanation of such features in terms of the opponent-process mechanisms of human visual systems.

**3.** Compare: "Water is my favorite beverage" may be interesting, but it does not tell us the kinds of things an acceptable scientific account of water must (viz., "Water is $H_2O$").



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

## Intrinsic changes in experience: Swift and enormous

Daniel C. Dennett

*Center for Cognitive Studies, Tufts University, Medford, MA 02155.*
**ddennett@tufts.edu** **ase.tufts.edu/cogstud/mainpage.htm**

**Abstract:** Because, as Palmer shows, the only kinds of differences that can be detected are differences in relational structure, and relational structure is precisely what is preserved by isomorphism, his own arguments can be used to expose the incoherent motivation behind the traditional idea of "intrinsic qualities" of experience.

As a left-handed person, I can wonder whether I am a left-hemisphere-dominant speaker or a right-hemisphere-dominant speaker or something mixed, and the only way I can learn the truth is by submitting myself to objective, "third-person" testing. I do not "have access to" this intimate fact about how my own mind does its work. It escapes all my attempts at introspective detection, and might, for all I know, shunt back and forth every few seconds without my being any the wiser. In striking contrast to this is the traditional idea that there are "intrinsic qualities" of my subjective experience that I do have access to, but that are inaccessible to objective investigation. This idea has persisted for centuries, in spite of its incoherence, but perhaps its days are finally numbered. Palmer presents the case in favor of the traditional view so clearly that his own arguments can be recast to expose the problems with it.

1. "The emerging picture is that the [intrinsic] nature of color experiences cannot be uniquely fixed by objective behavioral means, but their structural interrelations can be. This means that, logically speaking, *any* set of underlying experiences will do for color, provided the experiences relate to each other in the required way" (sect. 2.2, para. 6).

2. "[T]he only kinds of differences that can be detected *behaviorally* [my emphasis] are differences in relational structure, and relational structure is precisely what is preserved by isomorphism" (sect. 2.3, para. 4).

Behaviorally, as contrasted with what? Experientially, introspectively, first-personly. The idea is that what cannot be detected behaviorally might nevertheless be detected from the first-person-point-of-view, as one says:

3. "I alone have access to these experiences" (sect. 3.3, para. 4).

But (3) must be defended against the apparently unthinkable hypothesis that *not even I* "have access to" the intrinsic qualities of my very own experience. What on earth could this mean? It could mean that there were intrinsic qualities of my experience whose comings and goings were, like the spatial properties of my language-comprehension and production activities, beyond my direct ken. But this invites the obvious retort: then they would not be properties *of my experience*! Now what could that mean?

Palmer shows us, by plotting the path from between-subjects to within-subject experiments. The real and imaginary within-subject experiments he discusses all require a "memory comparison" by the subject. That is the whole point of within-subject experiments here, and Palmer acknowledges the theoretical possibility that there might be intrinsic qualities that changed so gradually, over such a long time, that the intrasubjective memory comparison would fail to detect them. If a change were slow enough, he concedes, even a huge change could occur without being detected, and if a change were subtle enough, it could happen quickly, without the subject noticing. But never mind, he says; he is concerned only with within-subject changes in experiential quality that are "swift and enormous" (sect. 3.5 para. 3). How swift and how enormous? Just swift and enormous enough *to be detected by the subject*.

Palmer concludes: "Within-subject designs can examine changes in experience, but cannot reveal to or from what they changed" (sect. 4 para. 9) – not to outsiders and not, really, to subjects either! You do *not* "have access to" the intrinsic qualities of your experiences in any interesting sense, any more than outside observers do, but only to the relations between them that you can detect. The very detectability *by the subject* of "swift and enormous" changes guarantees that any such changes of properties are "within the domain of functionalism." This does not establish that there are no "subisomorphic" intrinsic qualities of experience, but only that if there are, they are of no importance to psychology (or "phenomenology"), because their presence or absence makes no difference to the subjective state of the subject. In the limiting case, you could gradually become a "color zombie" and never know it. For all you know, that is what you are now. This is not, as some have claimed, an intended reductio ad absurdum of the very idea of consciousness, but rather of the idea that consciousness has intrinsic qualities that are problematic for functionalism.

Palmer notes that earlier arguments against intrinsic qualities of subjective experience (e.g., Dennett 1991) were "complex, convoluted, and generally unconvincing" (sect. 3.5, para. 2). Perhaps now that they are recast as implications of his own arguments, he and others will find them more persuasive.

## What does my eye tell your mind?

Rebecca M. Frumkina

*Department of Psycholinguistics, Institute of Linguistics, Russian Academy of Sciences, Moscow 125 167, Russia.* **frum@frum.mccme.ru**

**Abstract:** Palmer's suggestion that color might serve as an appropriate field in which to test the hypotheses on the relations among brain, mind, and behavior is misleading. Human color experience as related to mind must be considered as a cultural, as well as a natural phenomenon. It is important to avoid approaching facts of language as mere nomenclature; otherwise, concepts are reduced to percepts.

I found Palmer's target article highly controversial, although for rather uncommon reasons: I am used to discussion at either the theoretic or metatheoretic level (i.e., on epistemological grounds) or statements challenging certain experimental designs and/or results. Even when a general argument is stimulated by doubt about the validity of a whole paradigm to which a particular line of research belongs, I would expect a clear restatement of what seems invalid, and how one accounts for the consequences.

By all means, let the epistemological status of any statement in a contribution be transparent to the qualified reader. An opinion, however reasonable, should not be taken as the ultimate truth; an untestable hypothesis should not be presented as testable, and a model (e.g., of color space) should not be discussed as having its own ontology – be it Newton's, Hering's, or Munsell's.

It is misleading (though common) to suggest that color might serve as an appropriate field to test hypotheses on the relations among brain, mind, and behavior. As to what goes on in the brain when we discriminate and categorize colors, psychophysiologists know next to nothing (Dubois 1997; Missa 1993). We can register wavelengths and show experimentally that they evoke certain color terms as responses (Chapanis 1965). Does this lead us further than we are in the domain of categorization of odors, where we still cannot register any psychophysical correspondences, whatever they might be? I doubt it, as do Sahlins (1976) and Dubois et al. (1997, p. 27).

After 20 years of research in color as represented by the facts of natural language (Frumkina 1984; Frumkina & Mikhejev 1996), I conclude that human color experience and perception as related to the mind must be considered a cultural, as well as a natural phenomenon. Hence, the relation should be investigated by methods based on the observable behavior, be they questionnaires, mapping words to Munsell tables, categorization of color stimuli, or color-term usage by natural speakers.

We "see" with neither our eyes nor our brains. We see with our minds. We need our minds, not just our brains, to tell crimson from red. We might as well skip this difference if we do not be-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

long to the culture where this difference is encouraged. Many people are subject to special cultural conditions that might make certain differences – negligible under neutral conditions elsewhere – crucial and behaviorally important. It is widely known that textile experts can sort samples of black thread into 16 categories, and those who promote lipstick will highlight exceptionally the shades of red and pink. Such experts would be able to make us "see" the difference, too, but by addressing our minds, not our brains, let alone our eyes (Goodwin 1997). This brings us back to the problem of "basic" colors.

The discussion of "basicness" actually refers to color names and not to their denotata; that is, it does not refer directly to color samples represented via Munsell tables. Thus, we come across "derived" but basic (and presumably universal) color terms, such as *orange* and *goluboj*, interpreted as *red* and *yellow, white* and *blue,* correspondingly. I argue that *goluboj* should be considered culturally basic for Russian, because Russian native speakers cannot designate most blue eye color and the common color of sky without this term. The general "basicness" of *goluboj* has been widely discussed (Corbett & Davis 1997; Moss et al. 1990), but I wonder whether the term has any culturally important manifestation outside the Russian language area.

On the other hand, I have never found any argument about *orange* possibly not being basic in Russian. However, as long as the corresponding Russian term *oranjevyj* is available for the color of an orange, it obviously belongs culturally to the words borrowed rather late. *Oranjevyj* is quite infrequent (cf. Corbett & Davis 1997; MacLaury 1997); it does not make part of any collocations, nor does it serve as a typical attribute for any culturally relevant objects.

In sum: It is important to avoid approaching facts of language as mere nomenclature; otherwise, cognition is reduced to recognition, concepts to percepts, and culture to nature (Sahlins 1976).

## Empirical assessment of colour symmetries

Lewis D. Griffin

*Department of Optometry and Vision Sciences, Ashton University, Birmingham B4 7ET, England.* **l.d.griffin@aston.ac.uk**
**www.vs.aston.ac.uk/staff/lgriffin.html**

**Abstract:** The quality of potential symmetries of the similarity structure of the Basic Colour Terms has been assessed. The assessment was made on the basis of a database of similarity judgements, made by subjects in response to linguistically expressed questions. All potential symmetries can be statistically rejected, although the well-known and some novel interpretable symmetries are shown to be approximately correct.

To investigate possible symmetries of the colours, a database of colour similarity judgements was amassed using questionnaires. Each questionnaire consisted of 200 questions of the form "which is the more similar pair A and B or C and D?" where A–D were randomly drawn from the 11 Basic Colour Terms (BCTs). The questions were linguistic; no colour samples were used. Only subjects who assessed themselves as having normal colour vision and spoke English as their main language were used.

A total of 47,557 responses were collected from 194 subjects. Ignoring the order of colours within a pair and the ordering of the two pairs, there are 1,485 possible questions. So on average, each question had 32.0 responses. The questions elicited varying levels of agreement. For example, 33 versus 3 choose Purple and Black as more similar than Green and Black, whereas subjects split 17 versus 17 over Green and Red versus Brown and White. If agreement rate is defined so that it is 91.7% and 50.0%, respectively, for the previous two examples, the average agreement rate across all questions was 79.1%.

To appreciate what a symmetry within the database of responses would amount to, imagine data being collected in the following manner. Suppose questions are asked in the form "A and B or C and D?" and each questionnaire is accompanied by a key translating letters into BCTs. The existence of a symmetry could be investigated by having two cohorts of subjects complete questionnaires, with each cohort using a different key. If the judgements of the two cohorts were found to be statistically distinguishable, the proposed symmetry, encoded in the relation between the two keys, would be rejected. Fortunately, given the number of potential symmetries, this clumsy method of data collection is unnecessary; the assessment can be made by comparing the response database to a transformed version of itself.

To describe how potential symmetries are assessed, consider a concrete example: the swapping of Red and Orange, Yellow and Pink, and Purple and Brown. First, each of the 1,485 questions is assessed for whether it is affected by the permutation of colours. A question such as "Red and Yellow or Pink and Blue" is affected, whereas "Red and Orange or Blue and Grey" is not. Next, consider an affected question such as "Purple and Orange or Yellow and Green." This had response 4 versus 26. After permutation of colours, the question becomes "Brown and Red or Pink and Green," which had response 32 versus 0. $\chi^2$ tests are then used to measure the discrepancy, in a weak and in a strong sense, between the two response patterns. The strong measure is the score from a $\chi^2$ test that the response patterns are identical; the weak measure is the score from a $\chi^2$ test of the hypothesis that the majority answer is the same in both cases. The weak and the strong $\chi^2$ scores for all the affected questions are separately summed to give overall scores for the symmetry. Because different symmetries affect different numbers of questions, the $\chi^2$ scores of different symmetries have different numbers of degrees-of-freedom ($v$) and so cannot be directly compared. To allow comparison, $\chi^2$ scores are normalised according to

$$Sds = \frac{\chi^2 - v}{\sqrt{2v}}$$

this is valid given the size of $v$. A Sds (standard deviations) value greater than 1.64 is evidence that a proposed symmetry should be rejected.

Table 1 shows the 5 best symmetries of the Hering primaries and Table 2 shows a selection of symmetries of all 11 BCTs. Both tables are ordered by the size of the weak Sds measure. The tables also show for each symmetry the question most violated by the responses. The remaining column, "fractional agreement," allows the quality of the symmetry to be assessed. It is calculated as the expected rate of agreement between two subjects using different keys (related by the symmetry), as a fraction of the expected rate if they use the same key. As can be seen from the Sd score in both tables all symmetries can be rejected. However, the fractional agreement scores show that there are several good approximate symmetries. Some of these are interpretable in relation to the standard octahedron model of colour space.

The best symmetry for the Hering primaries can be pictured as a 180° rotation about the Red-Green axis. As shown by the "worst question" for this symmetry, its main flaw lies in the mapping of Blue-White to Yellow-Black. The second best symmetry is a reflection in the Black-White-Yellow-Blue plane. Its main flaw is the mapping of Blue-Red to Blue-Green. The next two symmetries do not preserve the topology of the colour octahedron. The fifth symmetry is the best that transforms all six colours. It is the composition of the first two symmetries, but can also be understood as a mapping of the primaries to their complements.

The best 9 symmetries of the 11 BCTs clearly violate the topology of the BCTs, for example, the first maps Blue-White to Black-White. They have good fractional agreements, however, in particular the third, which has the highest of all the potential symmetries. The best symmetry that is reasonably in accord with topology is the 10th, which corresponds to a one-step rotation of the hue circle. Perhaps its worst violation of topology is the mapping from Blue-Brown (not adjacent) to Purple-Brown (adjacent). The next topologically reasonable symmetry is the 12th, which can


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

Table 1 (Griffin). *The 5 best symmetries (out of 397) of the 6 Hering primaries*

| Symmetry number | Symmetry | Weak Sds | Strong Sds | Fractional Agreement | Worst Question |
|---|---|---|---|---|---|
| 1 | Yellow ↔ Blue, Black ↔ White | 7.3 | 25.9 | 91.7% | Y & B (4) vs. Bl & W (30)<br>Bl & W (30) vs. Y & B (4) |
| 2 | Red ↔ Green | 9.7 | 26.8 | 89.7% | Bl & G (26) vs. Bl & R (4)<br>Bl & R (4) vs. Bl & G (26) |
| 3 | → Red → Blue → Yellow →<br>Black → White → | 11.6 | 42.7 | 87.1% | Y & G (22) vs. G & R (8)<br>B & G (5) vs. G & Bl (26) |
| 4 | Red ↔ White | 11.9 | 31.2 | 89.1% | R & B (27) vs. W & B (3)<br>W & B (3) vs. R & B (27) |
| 5 | Yellow ↔ Blue,<br>Black ↔ White, Red ↔ Green | 12.6 | 48.6 | 85.3% | Y & B (4) vs. Bl & W (30)<br>Bl & W (30) vs. Y & B (4) |

Color code: Y = yellow; B = black; Bl = blue; W = white; R = red; G = green.

Table 2 (Griffin). *A selection of the 19,976,247 potential symmetries of the BCTs*

| Symmetry number | Symmetry | Weak Sds | Strong Sds | Fractional Agreement | Worst Question |
|---|---|---|---|---|---|
| 1 | Black ↔ Blue | 18.2 | 74.7 | 92.2% | Bl & W (29) vs. W & B (1)<br>B & W (1) vs. W & Bl (29) |
| 2 | Yellow ↔ Orange | 23.5 | 92.1 | 89.8% | O & W (0) vs. Y & W (30)<br>Y & W (30) vs. O & W (0) |
| 3 | Red ↔ Orange, Y ↔ Pink,<br>Green ↔ Purple | 24.2 | 92.1 | 93.0% | Pu & B (33) vs. G & B (3)<br>G & B (3) vs. Pu & B (33) |
| 10 | → Red → Orange → Yellow →<br>Green → Blue → Purple → | 30.8 | 109.6 | 90.8% | O & W (0) vs. Y & W (30)<br>Y & W (33) vs. G & W (1) |
| 12 | Red ↔ O, Y. ↔ Pu, P ↔ Br,<br>Green ↔ Blue, Black ↔ W | 32.3 | 103.6 | 91.4% | P & Bl (4) vs. Br & G (43)<br>Br & G (43) vs. P & Bl (4) |
| 19 | Y ↔ Blue, Black ↔ W,<br>Orange ↔ Pu, Pink ↔ Brown | 35.9 | 113.8 | 90.6% | P & G (2) vs. Br & Y (31)<br>Br & G (43) vs. P & Bl (4) |
| 48 | → R → Bl → Y → Br → Grey → P<br>→ Pu → Green → O → B → W → | 41.1 | 142.1 | 88.4% | Bl & Gr (27) vs. Gr & W (1)<br>Y & P (3) vs. O & R (30) |

Color code: Y = yellow; B = black; Bl = blue; W = white; Pu = purple; R = red; G = green; O = orange; P = pink; Br = brown; Gr = grey.

be understood as a 180° rotation about an axis through Red/Orange-Grey-Blue/Green. I am unaware of any previous mention of this approximate symmetry. The 19th symmetry is rotation about an axis through Red-Grey-Green: This is the first symmetry of Table 1. The 48th symmetry is the best that transforms all 11 BCTs. It neatly maps Black-Grey-White to White-Pink-Red, Red-Pink-White to Blue-Purple-Red, Orange-Brown Black to Black-Grey-White, and Red-Orange-Yellow-Green-Blue-Purple to Blue-Black-Brown-Orange-Yellow-Green.

## Color relations and the power of complexity

C. L. Hardin

*Department of Philosophy, Syracuse University, Syracuse, NY 12344-1170.*
**chardin1@twcny.rr.com**

**Abstract:** Color-order systems highlight certain features of color phenomenology while neglecting others. It is misleading to speak as if there were a single "psychological color space" that might be described by a rather simple formal structure. Criticisms of functionalism based on multiple realizations of a too-simple formal description of chromatic phenomenal relations thus miss the mark. It is quite implausible that a functional system representing the full complexity of human color phenomenology should be realizable by radically different qualitative states.

Color-order systems are convenient representations of selected aspects of color experience. Each one is designed to exhibit certain features and, inevitably, omit others. For example, the Munsell system and the Natural Color System (NCS) are both meant to represent the appearance of reflective samples. But whereas Munsell hue samples are meant to be equispaced, and the system gives no pride of place to the unique hues, NCS hue samples are spaced according to their degree of resemblance to the unique hues, but are not equispaced. A Munsell Chroma step is determined by an estimation of the gray content of a sample as compared with a gray of the same Munsell Value. An NCS chromaticness step is determined by an absolute estimation of the degree of chromatic content in a sample. The Munsell system is defined by the samples of its atlas. The NCS system is based on comparisons



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

with mental prototypes of the six Hering elementary colors (this is not a crazy idea!) and the NCS atlas is intended only as an illustration of the system. On the other hand, the HBS (hue-brightness-saturation) system is meant to map the relationships of colored lights; two of its dimensions do not map exactly onto the (roughly) corresponding dimensions of either Munsell or NCS.

So which of these systems (which by no means exhaust the field) represents psychological color space? All of them – and none of them. Not only does each lack an important feature of one of the others, many important relationships fail to be represented by any of them. For example, the strikingly unequal size of the green and red regions that are picked out by all languages that have green and red as basic color terms can be marked out in either the Munsell or NCS systems, but they are not represented by any structural features of those systems. (Those who suppose this size inequality to be a cultural artifact should read Matsuzawa [1985], who describes how a chimpanzee generalized from 11 focal color chips in much the same way as a human being.) Some of the similarities and dissimilarities that the Hering elementary colors seem to bear to each other are completely obscured by both of these systems. For example, there is some suggestive evidence that people see elementary green and elementary blue to be more like each other than are elementary red and yellow. (The question has not been systematically investigated. Any takers?) This matter is not directly resolvable by counting just-noticeable differences from one elementary color to another; it is an open question as to whether the sum of small color differences is an adequate measure of large color differences.

Although the models that are in general use exhibit some of the essential features of perceived color, it is misleading to speak of "*the* psychological color solid" as if there were a unitary and simple psychological model that captures the entire range of color phenomena as we experience them. This is one of several reasons why not even someone "in the grip of functionalism" (sect. 2.4) should have the slightest inclination to suppose that Palmer's "color machine" might perceive color. Is it "surprisingly difficult to prove that this machine fails to have color experiences"? It is perhaps equally difficult to *prove* that this stone is not now thinking of Vienna (*pace* Carnap). After all, could not the stone be thinking a monadic Viennese thought that it was totally unable to express? The absurdity of this case arises in part from the rock's inability to engage in intelligently guided behavior. A bare minimum criterion for something to perceive color is that color can be used to govern a varied behavioral repertoire. The one-trick pony that is Palmer's color machine would be far too impoverished to serve even as the color module for a robot that is to use color in its business of avoiding and identifying objects in a real world of occlusion, shadows, and variable illumination. Constructing something adequate to this task is not a trivial undertaking, as the robotics vision community will attest.

The existence of chromatic blindsight shows us that even when receptors are responding, and some degree of opponent processing is present, and some crude color naming can be elicited, there is no color experience. What is missing is, in a fashion, not mysterious. The peripheral color mechanisms are not properly connected to the central mechanisms. In another respect there is here a very deep mystery indeed, though not an intractable one: the puzzle of what those more central mechanisms are and how they interact with the rest of what is going on in the brain. But when the connections are properly made, the color-relevant mechanisms have become extremely complex, and so has the color-related behavioral repertoire that these mechanisms make possible. I invite you to spend an agreeable evening reading a sensitive account of the rich texture of chromatic relationships such as that in chapter three of Charles Riley's *Color codes* (1995). If you do so, I would venture to say that you will not find it even remotely possible that the qualities of color experience in Riley's brain are radically different from yours. As the level of behavioral complexity increases dramatically, the number of alternative models drops precipitously.

Analogous situations in the physical world are commonplace. Consider a bedspring that is compressed and then released. Its behavior is well described as a damped harmonic oscillator. Alternate models satisfying this formal structure include a radio circuit and a loudspeaker in an enclosure. Neither of these is particularly suitable for sleeping, any more than the Palmer machine is suitable for guiding animal behavior. But there is no mystery about distinguishing bedsprings from radios: just give a fuller description, and alternative models become scarce. Indeed, a sufficiently rich description of materials and behavior would uniquely single out a Simmons bedspring manufactured in 1959. Of course for any system of relationships, however rich, there could always be alternative models fashioned by Berkeley's God or Descartes' evil demon, but it could never be the task of scientific explanation to rule these out.

## Logical possibility and the isomorphism constraint

Bernard Harrison

*Department of Philosophy, University of Utah, Salt Lake City, UT 84112.*
**bernh@globalnet.co.uk**

**Abstract:** Palmer's "isomorphism constraint" presupposes the logical possibility of two qualitatively disparate sets of sensory experiences exhibiting the same relationships. Two arguments are presented to demonstrate that, because such a state of affairs cannot be coherently specified, its occurrence is not logically possible. The prospects for behavioral and biological science are better than Palmer suggests; those for functionalism are worse.

I seem to have been the first to explore at length (Harrison 1967; 1973) the idea that asymmetries of relative similarity between colors might prove fatal to philosophical scepticism based on Locke's inverted spectrum argument. However, I soon ceased to find the arguments I had devised by 1973 altogether satisfying. As I saw it, I had failed to deal adequately with the possibility, which Palmer makes the main plank of his argument for his "isomorphism constraint," that the color-experience of B might be qualitatively different from that of A (who perceives what we would call colors) in ways beyond A's power to imagine, yet be, nevertheless, structurally isomorphic with A's color-experience, in the sense of preserving all the relations obtaining between colors as we, and A, perceive them.

I therefore continued to think about the problem, and by 1984 had devised additional arguments, which appeared in print in a slightly abridged form two years later (Harrison 1986, pp. 112–14) and in a fuller version the year after that (Harrison 1987, pp. 184–87).

The arguments in question, which, because they are available elsewhere, I will do no more than summarize briefly here, are designed to show that the existence of qualitatively disparate yet "structurally" or "relationally" isomorphic color experiences is not a possibility, not even a "logical" one! The notion of "logical possibility" can be deeply misleading in philosophy, and has the potential to do even more damage in science. The problem is that it is quite easy to formulate "possibilities" that, while they look entirely plausible and above board at first sight, are nevertheless incoherently specified.

Such is the case, it seems to me, with the "possibility" of qualitatively different yet relationally isomorphic sets of color-experiences. To say that the two proposed sets of color-experiences are relationally isomorphic is presumably to say that the qualitatively discriminable sensory presentations making up each set exhibit *the same relationships* to one another as those making up the other set. But what is *the same* supposed to mean in this context? Palmer suggests, in effect (sect 2.3), that the required notion of sameness be explained by analogy with the identity of the sets of mathematical relationships linking the primitive elements of the alter-


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

native interpretations of a given axiom scheme constituting a *dual system* in mathematics. An important disanalogy between the two cases, however, is that whereas in the mathematical case the applicable notion of sameness can be specified without reference to the set of primitive elements constituting either interpretation by reference to the uninterpreted axiom set, no such move is available in the case of color. The "space" of color defined by relationships of relative similarity between color presentations can only be generated by continuous, qualitative modification of the "primitive elements" (the color presentations) composing it.

There is thus no way of characterizing the relationships defining such a space without reference to its primitive elements. If now, guided by notions of "logical" possibility better adapted to the needs of metaphysical speculation than to those of behavioral science, we attempt to envisage a space (B) of sensory presentations that would not *be* (would, indeed, be unimaginably different from) *color* presentations, but would nevertheless exhibit *the same relationships* as the space (A) of colors, we find that, in dispensing with the primitive elements of space (A) in favor of the (unimaginable) primitive elements of space (B), we have deprived ourselves of any means of characterizing, with respect to space (B), the relationships of relative similarity that define space (A). Put bluntly, the incoherence vitiating this kind of sceptical hypothesis comes to a head in the question of how, if the primitive elements of space (B) are not, qualitatively speaking, color presentations, space (B) could be defined by relationships of relative similarity, which we have no means whatsoever of characterizing other than as the relations generated by the possibilities of qualitative modification qualitatively inherent in color presentations?

Once one sees this, it is easy to show (Harrison 1986; 1987) that in an actual case of apparently abnormal vision, space (B) must either be the familiar space of color, and the apparent abnormality otherwise explicable, or else it must result from the operation of some sensory modality other than color, which will have its own relational structure. The apparent third possibility, of an alternative sensory modality having *the same relational structure* as color, thus requires no empirical work to exclude it, but can be dismissed out of hand as a philosophical chimaera for the reason just advanced, namely, that the notion of sameness required by such a supposition turns out on inspection not merely to defy formulation but to be actually internally incoherent.

If the consequences of putting this antique philosophical warhorse out to pasture are damaging to certain of Palmer's tactical moves, they seem entirely consistent with his objectives. The "isomorphism constraint" collapses, but so, by the same token, does the idea that behavioral science is not fully adequate to deal with "the nature of experience," and that behavioral methods, which, as Palmer shows, are perfectly adequate to characterize in terms of its relational characteristics, what color a given observer is perceiving, nevertheless remain somehow incapable of revealing "deeper," more "subjective" aspects of the subject's experience. The ultimate moral of the above argument, indeed, is that we need to reexamine carefully and suspiciously, perhaps along the lines suggested by Gaston Bachelard (1972), the intellectual credentials of the familiar distinction between "objectivity" and "subjectivity" according to which behavioral methods are "objective," whereas "qualia," "the content of consciousness," and so forth, are "subjective."

Finally, of course, inasmuch as functionalism derives its main theoretical justification from the plausibility of the distinction between the "subjective" and the "objective," considered as distinct "realms" or departments of reality, to which it presents itself as a response from the point of view of "materialism," undercutting the entire distinction between "materialism" and "subjectivity," by appeal to the broadly Bachelardian principles advocated here, will tend to undercut the need for functionalism, weakening its power to resist objections of the type tellingly advanced by Palmer.

## If not functionalism, then what? Eliminative materialism?

Harry Howard

*Department of Spanish and Portuguese, Tulane University, New Orleans, LA 70118.* **howard@mailhost.tcs.tulane.edu**
**www.tulane.edu/~howard/HHHome.html**

**Abstract:** The isomorphism between relational structures advocated by Palmer corresponds quite closely to Paul Churchland's theory of "state-space semantics," so much so that one can be used to elucidate problematic areas in the other. The resulting hybrid shows eliminative materialism to be superior to functionalism as a theory of mental phenomena and seems to provide the best ontology for cognitive science.

Palmer elaborates two arguments against the functionalist approach to mental phenomena: (1) humans can have different experiences with the same relational structure, and (2) artificial systems can be built that are causally isomorphic to humans but lack conscious experiences altogether.

Palmer devotes the bulk of his target article to the first point, in particular to the elucidation of the isomorphism between color percepts and the color solid of psychometric testing. It is helpful to define the relational structures in question. The first is a set of color percepts *C*, ordered by the three relations red-green, $\leq_{rg}$, blue-yellow, $\leq_{by}$, and light-dark $\leq_{ld}$, which correspond to the three parameters of color that human vision is sensitive to. The second is the three-dimensional color solid organized on three axes that can be abbreviated as $\leq_x$, $\leq_y$, and $\leq_z$. These ingredients are tabulated in (1):

| 1. | *Relational formalism* | *Realization in color science* |
|---|---|---|
| (a) | $K = <C, \leq_{rg}, \leq_{by}, \leq_{ld}>$ | output of V4, that is, color percepts |
| (b) | $\Sigma = <S, \leq_x, \leq_y, \leq_z>$ | psychometric color space, that is, the color solid |
| (c) | $K \leftrightarrow \Sigma$ | color psychometric isomorphism |

Palmer's main contention is that external observers can never know an element of K directly; they can only know the place that such an element maps onto in $\Sigma$. In other words, in principle, there is no way to know whether two people share the same elements of C, that is, whether they see the same color percepts. In practice, there may be a way if the relational structures K and $\Sigma$ are asymmetric enough so that there is no distortion of one that cannot be reflected in the other under isomorphy. Unfortunately, the color solid has at least two axes of symmetry, so a reversal of one such axis in K can still be mapped isomorphically into $\Sigma$ without detection.

As for the second argument, that artificial systems can be built that are causally isomorphic to human color perception yet lacking in any kind of color experience, I have nothing to add except that such systems are closer to realization than Palmer suspects (see the neural-network model of Courtney et al. 1995a; 1995b).

The conclusion is that functionalism cannot be correct, regarding the first argument, for it would imply that "normal" and "reversed" color experiences are equivalent, when they clearly are not. Likewise, regarding the second argument, it would imply that a neural network along the lines of the one designed by Courtney et al. would have the same color experiences as humans, whereas it would surely not.

I find both arguments compelling, and so would like to know where to turn for an alternative to functionalism, but this is as far as Palmer goes.

Fortunately, I have an inkling of my own of where to turn, namely, to the theory of eliminative materialism proposed by Churchland (1981) and elaborated on in Churchland (1986; 1989; and especially in 1995). In particular, Palmer's notion of isomorphisms between relational structures is subsumed by Churchland's "state-space semantics," to use the felicitous coinage of Fodor and Lepore (1992). A vector space with a suitable metric defined on it reproduces all the properties of Palmer's relational



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

structures, and coordinate transformations of vectors reproduce all the properties of Palmer's isomorphisms between relational structures. (2) gives a first approximation of the correspondence between the two formalisms, where the relation ≤ on the left corresponds to a measure of distance *d* on the right, and the transformation between the two structures is accomplished through multiplying by the matrix *M*:

2. *Relational formalism* *Vector-space formalism*
(a) $K = <C, \leq_{rg}, \leq_{by}, \leq_{ld}>$ $K = <C, d_{rg}, d_{by}, d_{ld}>$
(b) $\Sigma = <S, \leq_{x}, \leq_{y}, \leq_{z}>$ $\Sigma = <S, d_{x}, d_{y}, d_{z}>$
(c) $K \leftrightarrow \Sigma$ $\Sigma = M * K$

The vector-space representations have considerable neurophysiological plausibility, as Churchland goes to great lengths to demonstrate.

An additional benefit of adopting eliminative materialism is that certain points that are obscure in one treatment are clarified in the other. For example, Churchland's (1995, p. 198) "proprietary, first-person epistemological access to some phenomenon," which is ultimately refined down to an "auto-connected way of knowing," is immediately recognizable as Palmer's subjectivity barrier. Churchland develops these notions to defend eliminative materialism from Nagel's (1974) bat and Jackson's (1982) neuroscientist. In a nutshell, the argument concerning Nagel's bat is that, given how different a bat's sensory experience is from a human's, no human will ever truly understand what it is like to be a bat. Palmer would say that this is an expected consequence of the subjectivity barrier: The best we can hope to do is find some kind of partial mapping between the relational structures of bat and human experience.

This cuts both ways, though, because flaws in one framework can adhere to the other. For example, Fodor and Lepore (1992, pp. 197–98) accuse state-space semantics of "pernicious holism," as explained forthwith. For Churchland, at any moment one's consciousness consists of a huge vector composed of all the active sensory modalities. Modalities are distinguished from one another by their position in the "consciousness" vector, whereas specific percepts within a modality are distinguished by the values of the vector at these positions. Fodor and Lepore (1992, pp. 197–98) object that under such a holistic approach, it becomes impossible to tell whether two individuals have the same concept. Concepts are to be located in this "consciousness" vector, yet the number or range of the dimensions of the spaces within this vector will vary from individual to individual and will therefore be incommensurate. If we cannot compare dimensions of representation between individuals, then we cannot compare the contents of the dimensions. Because Palmer claims that relational structures are necessary both to distinguish between different modalities of perception and to distinguish specific percepts within a modality, and Palmer's relational structures can be correlated with Churchland's state spaces, it follows that the charge of pernicious holism also sticks to Palmer's framework.

Fortunately for Palmer and Churchland, Laakso and Cottrell (1998) have found a technique to neutralize this objection. They show that a two-step algorithm can reliably measure the similarity of representations embedded in different spaces. The first step is to measure the distances between like percepts in each space, and the second step is to calculate the correlation between these distances across spaces. [See also Edelman: "Representation Is Representation of Similarities" *BBS* 21(4) 1998.] Note the resurfacing of Palmer's subjectivity barrier: We cannot compare percepts themselves, but only their correlates within the encompassing relational structure. So Churchland's, and by extension, Palmer's, account is vindicated, and we may be a step closer to a global theory of cognitive science.

## Disorder of colour consciousness: The view from neuropsychology

Glyn W. Humphreys and M. Jane Riddoch

*Cognitive Science Research Centre, School of Psychology, University of Birmingham, Birmingham, B15 2TT, United Kingdom.*
**g.w.humphreys@bham.ac.uk**

**Abstract:** We discuss the difficulty of measuring the perceptual experience of colour, supporting Palmer's assertion that neuropsychological disorders of colour processing can be informative in this respect. We point out that some disorders seem to affect the perceptual experience of colour over and above the perceptual processing of colour, providing direct insights into the neural mechanisms supporting perceptual experience.

In his target article Palmer points out the difficulty of analyzing the perceptual experience of colour through objective behavioural means. He points out the problems involved in assessing the quality of a perceptual experience of colour, particularly by means of between-subject analyses. Like Palmer, we are sceptical about even the best hopes for this approach. For example, attempts to define neurological equivalence classes of colour experience, based on common neural substrates of colour-related brain activity across individuals, still run into the subjectivity problem, here in knowing whether the experience of colour is the same even when brain states are. Any argument for equivalent experiences remains a leap of faith, and one that requires a belief that there are few individual differences in neural localisation.

A solution to this dilemma, Palmer suggests, is to conduct within-subject analyses. An example of this is acquired cerebral achromatopsia, where a patient may become functionally colour blind following a brain lesion. Such patients can detect a change in their colour experience after their lesion, indicating that measurements of colour experience are possible within individuals, and that these can also be related back to underlying physiology (here in terms of the site of any lesion). We wish to point out that studies of such patients can be even more informative than this, because colour experience itself can be affected, over and above effects on colour processing. Humphreys and colleagues (1992) reported data on one such patient, HJA. Like many achromatopsic patients, HJA suffered bilateral damage to occipito-temporal regions, including the lingual and fusiform gyri, and this not only led to problems in colour perception but to difficulties in object and face recognition. Though he had formerly worked in fine art and advised on colour, HJA reported that he saw the world only in terms of shades of grey following his lesion. He detected a change in his colour experience. Nevertheless, aspects of colour processing remained. For example, visual evoked responses to isoluminant colours could be measured and his ability to match isoluminant colour patches was above chance. HJA's conscious reports were clearly impaired, however, when he was asked to judge whether he was right or wrong when making colour-based responses (and even when the colour responses were correct). Thus, when asked to point to a colour token matching one pointed to by the examiner, HJA's confidence judgements bore no relation to the accuracy of his performance. He often felt he was guessing when he was right and he felt confident of being right when he was in fact wrong. His conscious experience of colour appeared to be dissociated from the residual colour processing abilities he had.

This dissociation of colour experience was also distinct from HJA's conscious experience of other perceptual impairments. Thus his judgements were generally accurate when he was asked to rate his confidence about whether object and face identification responses were correct. He also showed no evidence of residual access to object or face identities, unlike the results with colour. Hence, in this instance, the degree of perceptual deficit – measured in terms of residual perceptual abilities – can be distinguished from the conscious experience of the deficit. For one class of stimulus (colour) there was better residual processing, but less insight into the deficit, than for other classes of stimulus (objects,


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

faces). It is not simply that patients with a more profound perceptual deficit experience a more profound loss of the ability. From this we may conclude several things. For example, it may be possible to distinguish the neural substrates involved in perceptual processing from those involved in conscious awareness of their products. The neural substrate of conscious experience may also take a distributed form, and so can be dissociated for different stimuli. We suggest that detailed analysis of such patients can inform us not only about perceptual processes, but about how such processes are realised in subjective experience.

## Overlooking the resources of functionalism?

Zoltán Jakab

*Institute of Interdisciplinary Studies, Carleton University, K1S 5B6 Ottawa, Ontario, Canada.* **zjakab@ccs.carleton.ca**

**Abstract:** Although the author's critical view of functionalism has a considerable intuitive pull, his argument based on the color room scenario does not work. Functionalism and other relational views of the mind are capable of providing coherent accounts of conscious experience that meet the challenge set up by the "color room argument." A simple example of such an account is presented.

Palmer claims that because functionalism can give only a relational picture of the mind, it will fail to capture the intrinsic qualities of experience. Experiential qualities are below the level of relational isomorphism that can be captured by the methods of behavioral science in general and functionalism in particular. In support of this claim, Palmer sets up an argument in two versions (sect. 2.5 and sect. 4). First let me reconstruct this argument.

***Version one.*** The color machine in the color room satisfies all functionalist (computational) requirements associated with color discrimination and color-related behavior. A functionalist should therefore conclude that the color machine in the color room has color experiences. But it is intuitively implausible that the color machine has any color experience. Therefore functionalism is probably wrong.

***Version two.*** Put yourself in the color room, thereby bypassing the other mind's problem. Master the computation that the color machine performs; in this case, you become the color machine, you satisfy all functionalist requirements for having color experiences, hence (so the functionalist must argue) you will have color experiences simply by means of doing that calculation. But again, very plausibly, you will not have any color experiences by means of doing that calculation. Therefore functionalism is presumably false.

Overall, the structure of the argument is modus tollens:

1. If functionalism is right, then the color machine (or you) necessarily have color experiences merely by means of performing the color-related computations.

2. Neither the color machine, nor you (merely by means of doing the relevant calculations) would have any color experience. Therefore

3. Functionalism is wrong.

Now let me give a reply. Functionalism might be wrong (i.e., incapable of accounting for conscious experience), but Palmer's argument based on the color-room scenario is insufficient to show this. The argument is not sound because the first premise is unsupported. (I will not address the second premise in this commentary, even though doubts might arise about it as well.) Let us see what the problem is with the first premise.

***Version one.*** The analogy between the *human brain as a whole* (or some implementation of its functionally/computationally relevant structure) and the color machine does not hold up. The color machine is at most the model of an isolated subsystem of the brain. Should the functionalist conclude that it has color experiences? I think functionalism is not at all committed to drawing this inference.

Compare: Would a visual brain *in itself,* isolated from the rest of the brain, floating in some suitable solution, receiving appropriate optical stimuli, have color experiences? *When embedded in the neural/functional architecture of the rest of the brain,* the well-functioning visual brain does give rise to color experience. But does it do so in isolation? This is questionable, to say the least.

***Version two.*** Does functionalism entail that the human subject in the color room must have color experiences simply in virtue of performing the relevant calculations? I think not. Here is why.

There is a possible analysis of experiential qualities, which is (1) relational and (2) not yet ruled out as insufficient: Perhaps the experience of seeing red is a relation between a subject and a certain type of physiological state. The relation is "undergoing" a state: a token of a physiological state type occurs in one's brain *in the appropriate way* – for example, it is a well-characterized activity pattern *of* area V4. (Additional background conditions like normal awake state or REM sleep, sufficient attention to events of visual perception, etc., can be assumed.) My experience of seeing red *is* the undergoing relation between me and that particular state – the relation set up by that physiological state occurring in my visual brain. Moreover, in describing such a physiological state, we necessarily resort to some kind of abstraction. This already happens when we specify the physiological activity *type* that, for example, is tokened in V4 when the subject sees red. In giving such types we leave out idiosyncratic biochemical and physiological variations as irrelevant and specify a generalizable physiological pattern. Perhaps we can even specify some sort of computational operation that is performed by that physiological event. Furthermore, perhaps seeing a color is a subject's undergoing relation to an inner state *type – identified at the computational level.* If this is right, then we have a functionalist account of qualia.

Intuitively, this account may seem too austere. Is this so much the worse for the account, *or* so much the worse for the intuition? A difficult question. It has not yet been convincingly argued by anyone (to my knowledge) that this account cannot be right. An obvious problem with it is that it has yet to be spelled out in reasonable detail; however, even in this extremely sketchy form it absolves the functionalist of the burden of concluding that the subject in the color room has to have color experiences simply by virtue of performing the color-related calculations. If I realize the color-discrimination process by calculating the appropriate computational algorithm in my head, this process will primarily involve my higher order cognitive machinery. Hence this mental simulation does not at all imply that the physiological states underlying color experiences occur in my visual system; mental calculation need not involve activity in V4 or any other color-processing area. If I am the person in the color room, then it should not be at all surprising that I have no color experiences merely by means of doing that calculation.

Note also that the proposed relational account predicts that zombies physiologically identical to us are logically impossible. Once human subjects entertain some relevant physiological state, then by definition they have the corresponding experience.

## Asymmetries in the distribution of composite and derived basic color categories

Paul Kay

*Department of Linguistics, University of California, Berkeley, Berkeley, CA 94720.* **kay@cogsci.berkeley.edu** **www.icsi.berkeley.edu/~kay**

**Abstract:** PURPLE (RED-and-BLUE) is the most frequently occurring derived (binary) basic color term (BCT), but there is never a named composite BCT meaning RED-or-BLUE. GREEN-or-BLUE is the most frequently named composite color category, but there is never a BCT for the corresponding derived (binary) category CYAN (BLUE-and-GREEN). Why?



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

Palmer notes that the naming of some but not all of the possible composite and derived basic color categories constitutes a puzzling set of asymmetries in color experience, assuming these categories are in fact reflections of color experience. He continues; "I am not aware, however, of any behavioral data that directly support these asymmetries for derived and composite color categories in color experience" (sect. 1.4). I think such behavioral data exist, though they are admittedly few. I return to this point presently. Meanwhile, I would like to flesh out further the question of which possible composite and derived categories actually occur, what their relative frequencies are, and what significance may be attached to these relative frequencies. I will restrict my attention to the 110 languages of the World Color Survey (WCS; see Kay & Berlin 1997).

Figure 1 summarizes much – but not all – of the WCS data on composite and derived categories. It also displays the number of just noticeable differences (jnds) separating each pair of nonopponent hue primaries (RED, YELLOW), (YELLOW, GREEN), (GREEN, BLUE), and (BLUE, RED). Palmer raises the issue of the relevance of jnd separations: "For example, if for some reason there are more just noticeable differences (jnds) between unique red and unique yellow than between unique green and unique blue, the wider psychological gap might explain why there are BCTs for ORANGE in many languages but not for CYAN (blue-green)" (sect. 1.4). Although I am not certain that the number of jnds separating two unique hues measures a psychologically real distance, this information has been included in Figure 1 to allow its possible relation to variations in the popularity of corresponding composite and derived categories to be assessed.

The immediately striking facts observable in Figure 1 are (1) that the most popular composite category, B-or-G, is composed of primaries whose corresponding derived category, CYAN (B-and-G), never receives a BCT, and (2) that the most popular derived category R-and-B (PURPLE) combines primaries whose potential composite category, R-or-B, never receives a BCT. One is tempted to speculate that BLUE and GREEN are in some yet-to-be-determined sense "similar." If similarity promotes the recognition of a composite category and dissimilarity promotes the recognition of an intervening, derived category, we have an explanation why BLUE and GREEN are readily lumped into a composite BCT and never separated by a derived BCT. If BLUE and RED are dissimilar, the same logic explains why BLUE and RED are most frequently separated by a derived BCT and never lumped into a composite.

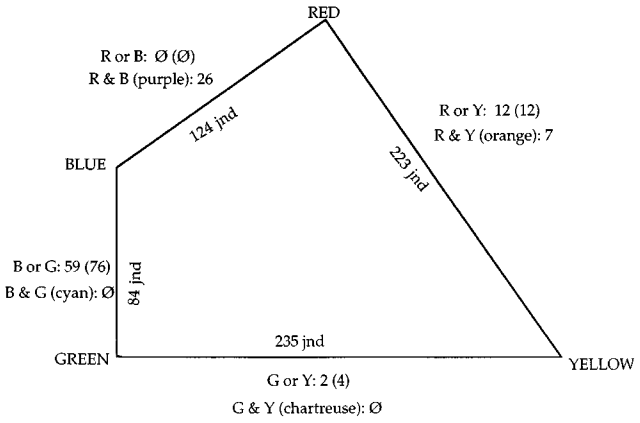



Figure 1 (Kay). Frequencies among 110 WCS languages of binary composite and derived categories, with the number of jnds separating the two primary categories of which they are composed. The first number given for a composite category records the frequency of that category. The second number (in parentheses) records the frequency of all composite categories containing the two indicated primaries. For example, 59 of the 110 WCS languages contain a B-or-G category; 76 languages contain either a B-or-G, or a B-or-G-or-Bk, or a B-or-G-or-Y category. (Source for jnd separations: MacLaury 1997b, p. 88.)

Unfortunately, neither the behavior of the remaining two pairs of non-opponent hue primaries nor the jnd separation data support the similarity conjecture. With respect to the pair (RED, YELLOW), there are 12 instances of the corresponding composite, R-or-Y, and 7 instances of the corresponding derived category (R-and-Y = ORANGE). With the pair (GREEN, YELLOW) we have a tiny number of composites and no instances of the corresponding derived category (CHARTREUSE). So although it might be attractive to speculate that nonoccurrence of CYAN as a BCT coupled with the great popularity of B-or-G makes sense in terms of the "similarity" of BLUE and GREEN and that the "dissimilarity" of BLUE and RED accounts analogously for the popularity of PURPLE and the nonoccurrence of a B-or-R composite, the hypothetical negative correlation between the number of composite and derived categories, for pairs of primaries, breaks down with the (GREEN, YELLOW) and (YELLOW, RED) pairs. Both the (GREEN, YELLOW) pair and the (YELLOW, RED) pair show a slightly higher incidence of the composite than the derived category, with inconclusively small absolute numbers.

With regard to jnd separation, the comparison of only the (BLUE, GREEN) and (RED, BLUE) pairs, 84 jnds and 124 jnds, respectively, appears to support the similarity hypothesis. BLUE and GREEN are indeed separated by fewer jnds than are RED and BLUE. But looking now at the jnd data for the other two pairs, (RED, YELLOW) and (GREEN, YELLOW), we note both of these jnd separations far exceed those of either of the first two pairs, destroying any hope of jnd separation being able to explain the relative popularity of composite and/or derived categories among different pairs of nonopponent hue primaries.

We have not been able to explain the striking fact that in the WCS data the most frequent two-primary composite BCT, G-or-B, corresponds to a derived category, (G-and-B = CYAN), which never gets a BCT and the most frequent derived BCT, (B-and-R = PURPLE), corresponds to a composite, B-or-R, which is never accorded a BCT. When the 98 languages of the Berlin and Kay (1969) study are added to the 110 languages of the WCS, this observation still holds. The negative correlation between the likelihood of a pair of non-opponent hue primaries being combined in a composite and in a derived category is confirmed in a large number of languages, but only in the extreme cases of the most popular composite and derived categories.

Returning to the question of whether there are behavioral data supporting asymmetries regarding composite colors, two experiments carried out by James Boster (1986) have been summarized as follows:

> The strength of the association of warm hues with W[HITE] and of cool hues with Bk is reinforced by experiments performed by James Boster (1986). In one experiment Boster gave twenty-one naive English-speaking subjects eight color chips, representing focal examples of the categories black, white, red, orange, yellow, green, blue and purple. The initial instruction was to sort the chips into two groups "on the basis of which colors you think are most similar to each other . . ." (Boster 1986: 64). The overwhelming preference was to put white, red, orange and yellow into one group and green, blue and black and purple into the other. Two-thirds of Boster's subjects chose this exact division into two subsets. (There are 2,080 ways a set of eight elements can be divided into two non-empty subsets.) In a second experiment, the same instruction was given to a group of eighteen subjects, using as stimuli the eight color words rather than the colored chips. Substantially the same result was obtained (Kay & Maffi, in press).

In subsequent sorts, Boster's subjects also showed strong preferences for keeping together the pairs (RED, YELLOW) and (GREEN, BLUE), the latter more strongly than the former.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

## The inverted colour space of vampires

Karel Kranda

*Psychophysiological Laboratory, Universitats Krankenhaus Benjamin Franklin, D-14057 Berlin, Germany.* **ksk@zedat.fu-berlin.de**

**Abstract:** Palmer's attempt to dust off Locke's construct of "inverted spectrum" is discussed here to examine its plausibility. Perceptual inversion could be fulfilled by adopting the notion of "inverted trichromacy" rather than by the proposed existence of "red-green reversed trichromats." Although the former alternative conforms to a hypothetical world of vampires, it fails to conform to the realities of genetics and neuroscience.

A compelling sequel to John Locke's (1690) discourse on colour, which Palmer alternately describes as "inverted spectrum argument" or "problem," may be the notion of "inverted trichromats," because only such radical mutants may appreciate the inherent beauty of an "inverted" rainbow. I would like to argue here that unlike in Locke's days, inverted trichromats no longer belong to the realm of philosophical constructs; their existence has long been proven beyond any reasonable doubt. Numerous publications and films meticulously document the lifestyle of these mutants, who are known as vampires and descend from a single man named Dracula. Suffering from photophobia, vampires go out at night while spending most of their daytime in crypts. Night is day to vampires and they, like most of us normals, prefer "light" to darkness. In the vampiric world, the moon shines "brighter" than the sun. But you would not notice this anomaly if you conversed with an average vampire because the semantic term "darker" would actually mean "brighter" to him. Vampiric sensation of colour is also totally inverted so blood appears "green" to a vampire and this apparent "greenness" has a soothing effect on his *psyche*.

In this dark light of the vampire world, Palmer's treatise is not particularly enlightening. Instead of proposing vampires as the true standard-bearers of "inverted trichromacy" – for who else but they can fully satisfy the conditions of the "complete inversion" set in his Figure 3c – Palmer fobs us off with some half-breed of "red-green–reversed perceivers" (sect. 1.3). These individuals supposedly exist (see Nida-Rümelin 1996) but are they more numerous than vampires?

The choice of an inverted spectrum as a paradigm for any discourse on colour is not a fortunate one, because it is easier to imagine it than to conceive any visual mechanism capable of such a feat. Should we ever discover such a "lucky," or on second thought, unlucky mutant, who sees the world in the "Lockian way," would we not face the intractable problem of locating the mechanism responsible for inverting his colour percept? The photoreceptor level is a poor candidate because any mirror-like inversion requires an even number of elements. But there are only three cone types at the retina. In theory, we can "reverse" the outputs of long- (L-) and middle-wave–sensitive (M-) cones but what about the short-wave–sensitive (S-) cones? The hypothetical construct of red-green reversal would not reverse the spectrum but at most only the percept within the red-green spectral range. Whether this reversal would preserve the asymmetry in the (L-M) opponent channel (Kranda & King-Smith 1979) is impossible to predict.

The mechanism proposed by Palmer as an explanation of "red-green reversal" is unusual as neither protanopes nor deuteranopes have "special" genes for "colour blindness." My interpretation of Palmer's argument is that he is implying a possibility of reversing the positions of the L- and M-pigment genes (LMG and LPG) in the X-chromosome array. The LMG is actually located upstream of the one or more MPGs (Vollrath et al. 1988). Thus in the "red-green reversal" case, the LMG would be located downstream, whereas the MPG would assume the upstream location. Each of these genes located in tandem has 6 exons, but only exons 2, 3, 4, and 5 differ from each other. The hypothetical case may thus require a complete reversal of those four exons at the two gene locations. The probability of any single exon mutating can be estimated from the frequency of inherited colour anomalies. Assuming the probability of exon mutation at about 0.03, the chance of the 2 sets of 4 exons actually reversing is no higher than 1 in $1.5 \times 10^{12}$. But even in this improbable "mutant," the "reversed gene," expressing the pigment type will also have to be expressed in the membrane structure of his cones, because the projecting dendrites of bipolar cells, like wine connoisseurs, cannot sniff the wine inside and must rely on the label. As there is no chance here of getting red wine in *vinho-verde* bottles, this "mutant" should preserve normal colour sensation.

Any type of "inverted trichromacy" expressing an inverted post-receptor connectivity presents the insurmountable problem of explaining its evolutionary origin. Unless we assume a macromutation going far beyond the hypothetical saltations of Gould (1980), it is difficult to conceive how the post-photoreceptor connectivity pattern could be reversed in one stride. Connectivity responsible for signalling colour and luminance, driven by mutations of genes specifying retinal pigments, must have taken millions of years to achieve full trichromatic performance. As only a gradual process can plausibly invert the connectivity pattern, the perceptual inversion would have to cross the centre of Palmer's colour space depicted in Figure 2. This may become a rather discomfiting experience for whole generations of mutants because this point equals zero and zero equals blindness.

Finally, it may indeed be "surprisingly difficult to prove that this machine [i.e., Palmer's own invention depicted in his Fig. 6] fails to have colour experience" (sect. 2.4), but probably not more difficult than proving the same claim made about a toaster. My guess is that if anyone ever decided to build Palmer's "colour machine" according to the "blueprint" provided, the finished product would achieve the colour experience of an average toaster. Unless the inventor withheld from us some important information, such as weighting functions specifying the valence of interconnections, for any given set of variables at the nodes "S," "M," and "L," the output values at the nodes "H," "B," and "S" would always equal zero.

## Isomorphisms and subjective colors

Gregory R. Lockhead and Scott A. Huettel

*Department of Psychology: Experimental, Duke University, Durham, NC 27708.* **{greg; huettel}@psych.duke.edu**

**Abstract:** Palmer describes a "subjective barrier" that limits knowledge of others' experience. We discuss how this barrier extends to all knowledge, becoming less distinct as theoretical constructs are strengthened. We provide evidence for isomorphic experience, among individuals with similar physiologies, by showing that perceived relations between colors are as similar when viewing pigments as when viewing subjective colors caused by flickering bars.

Palmer grapples with two questions that are fundamental to the philosophy of mind: How do I know that my perceptions are similar to yours, and how might we explain subjective experience? The first question is that of transformed qualia, that you and I may see the same object or color but might not have the same conscious experience. This may not be a productive scientific question because, as Palmer observes in Note 17, showing that two people's perceptions are the same requires accepting a null hypothesis. Concerning what evidence is sufficient for an explanation, we agree with Palmer that relations between elements determine the properties of a set, that isomorphic relations between set structures indicate association, and that correlated isomorphic structures do not imply an explanation, even though they might well reflect an explanation. As the idea of a "subjective barrier" describes, it is often difficult to know which is the case.

The barrier between correlational and causal explanations of phenomena is not unique to color or other psychophysical questions. It holds for all of knowledge. Consider I = E/R in elec-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

tronics. When we first learned this in school, it was difficult to understand how the mapping of resistance onto current is explanation rather than only correlation, but our teachers had no such difficulty. That is because we did not have the considerable theory and additional evidence they had. Learning about hydraulic flow made it easier to understand, or at least to predict, current flow, so we accept I = E/R as explanation. The subjective barrier is only considered to be crossed when we have strong theoretical constructs and converging data sets that link the subjective to the objective. It still exists as a logical construct but it becomes less troubling as data cumulates.

We may never gain enough knowledge about color to ignore the barrier, but we should come closer as more and different observations are added. Toward this end we want to add observations that might help move away from the concern that people who have similar physiologies and histories perceive identical things differently.

***Subjective colors.*** Subjective colors are seen when various black-and-white patterns are flashed or rotated (Cohen & Gordon 1949). Shown at the upper right of Figure 1 is a half-black, half-white sheet of paper (wrapped around a cylinder) with black lines in the white portion. When this cylinder is rotated at about 7 revolutions per second, the black lines appear colored. (White et al. 1977).

Color-normal people matched the flickering colors seen in this display to patches of colored papers (Munsell chips). They reported that subjective colors are more saturated when the lines are longer, that the dominant wavelength decreases as the distance between the black/white boundary and the line increases, and that the color structure reverses when the cylinder is rotated in the opposite direction. (Perhaps a reader can explain why distance-of-line-from-boundary is isomorphic with wavelength; it is unlikely this is a chance outcome and its basis is not understood.)

When color-normal people rate these flickering lines for similarity, the results are consistent with their similarity scaling of colored patches of paper and with Newton's color circle – very different wavelength equivalents (such as flickering bars reported as red and blue, or about 600 and 450 nm) are more similar to each other than either is to intermediate wavelength equivalents (such as green, or about 530 nm). This means that the physically one-dimensional ordering of wavelength that correlate essentially perfectly with border distance are transformed by sensation or perception into a two-dimensional representation. Ostensibly, this occurs because opponent-process coding in the retina maps the physically linear space of wavelength onto a space where physically different reds and purples are close to one another, and where the physically more similar reds and greens are more different from one another. The structures of the physical stimuli and the perceptual representations are not isomorphic, but the latter is a transformation of the former.

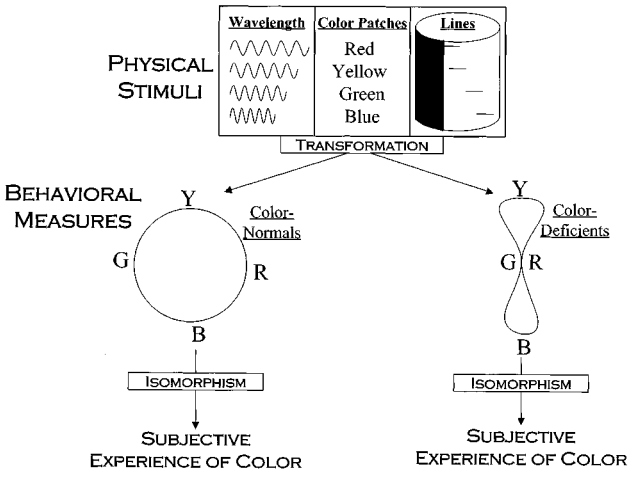



Figure 1 (Lockhead & Huettel). The structure of color perception in color-normals and color-deficients. For both observer types, perception of colors resulting from reflected wavelengths is similar to perception of subjective colors resulting from flickering lines. However, there are differences across observer types, as seen in the similarity scaling solutions.

Color-deficient observers (protanopes and deuteranopes) were also asked to scale the flickering bars for similarity, and the resulting similarity space is different from that obtained from color-normals. Rather than requiring a two-dimensional description (red-green and yellow-blue axes), a one-dimensional scaling solution seems to be sufficient; it accounts for 93% of the variance in the data (White et al. 1977). This is isomorphic with the results obtained when color-anomalous people scale Munsell chips for similarity (Shepard & Cooper 1975). Just as when color-deficient people judge colored papers, judgments of subjective colors show that Newton's color-circle (based on color-normal observers) appears as if a string has been tied around it at the red-green positions and pulled taut, resulting in an hourglass or a linear shape.

Similarity spaces for colored paints and flickering lines are isomorphic for color-deficient observers, and are isomorphic for color-normal observers, but these spaces for the different observer types are not isomorphic with each other (see Fig. 1). The perceptual structures match within groups but not across groups. From this, it is reasonable to conclude that color-normal people see subjective colors and colored paints similarly, because the two similarity structures are isomorphic, and they also see these colors identically, because the colors from one set are selected as equivalent to those from the other set. It is also reasonable to conclude that people with the same visual anomaly see colored paints and flickering lines equivalently, and that their perceptions are not the same as those of color normals.

These observations extend Palmer's argument that perception of color is isomorphic among individuals to observer type and to flickering lines. It also allows the suggestion that color perception must be explained at the level of similarity measures because, for both observer types, these are isomorphic with perceptual reports of standard colors and of subjective colors, but not with physical measures of either stimulus set.

## Asymmetry among Hering primaries thwarts the Inverted spectrum argument

Robert E. MacLaury

*Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19104.* **maclaury@sas.upenn.edu** **www.sas.upenn.edu/~maclaury/**

**Abstract:** Purest points of Hering's six primary colors reside at different levels of lightness such that inversion of each hue pair would be detectable in subjects' choice of foci on the Munsell array. An inverted spectrum would not impose the isomorphism constraint on a contrast of red-green or yellow-blue, whatever we conclude about inference in functionalism.

Palmer and I agree that both composite and derived basic color categories are asymmetrical (target article, sect. 4.1; MacLaury 1997b), which endows them with a structure capable of revealing inversion. But I disagree that primary basic white, black, red, green, yellow, and blue are symmetrical and, thus, that inversion of them would be undetectable (sect. 2.3). Figure 1 diagrams my reason for objecting. Each part, A or B, depicts the ethnographic array of 330 Munsell color chips. The left column delineates the achromatic continuum between white and black poles. The other 40 column delineates the achromatic continuum between white and black poles. The other 40 columns represent hue at the maximal saturation attainable by Munsell chips, which range across the rainbow from left to right and downward from light to dark (See MacLaury 1997c for specification). Each row-column inter-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

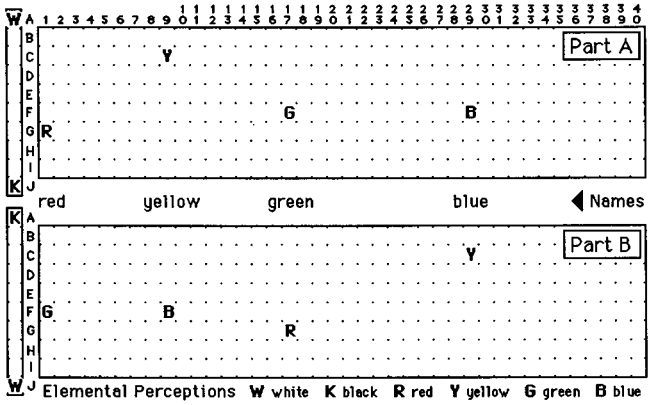



Figure 1 (MacLaury). Elemental points among Hering primary colors, (a) ethnographic pluralities of foci (MacLaury 1997b, p. 202) and (b) inversion of pairs W-K, R-G, and Y-B to K-W, G-R, and B-Y.

section plots a separate chip of distinctive pigmentation. Between parts A and B, four hue names are each affixed to a separate vertical band of the spectral array above and below the names. "White" and "black" are not shown, but they pertain, respectively, to rows A and J (upper and lower left column) where they, too, would be affixed.

Part A extracts information from MacLaury (1997c, p. 202, Fig. 1), wherein I compile 15,186 foci (best-example selections) of all color terms collected by the World Color Survey from 2,476 speakers of 107 minor and tribal languages in response to the ships. Marked as W, K, R, G, Y, and B are the six densest clusters of foci on noncontiguous chips, a worldwide vote for the closest approximation to the six elemental color points. As shown in MacLaury (1997c), these attract only pluralities of focus selections, not the majority. But the rest decrease in frequency proportionally to their distance from the six peaks, forming a histogram of stepped troughs between the apexes. If we were to guess in advance of an ethnographic interview where the subject would focus his or her primary basic color terms, chips A, J, G1, C9, G17, and G29 would offer the best chance of accuracy. There are many reasons why individuals place foci in the troughs, including genetically determined variation in hue perception (Neitz & Neitz 1995) or especially cognitive processes (MacLaury 1997b). But part A is likely to depict within a close tolerance the pure perceptions of normal trichromats everywhere.

Part A reveals a relational structure among the universal pure points, with W and K at extremes of lightness, R slightly but detectably darker than G, and Y markedly lighter than B. We must predict of any individual that his or her primary basic color categories will be identified with this structure and will be focused in this relation to each other. Part B shows the same foci after Hering color pairs have been inverted: W-K to K-W, R-G to G-R, and Y-B to B-Y. The inversions within the latter two pairs are betrayed by the asymmetrical lightness levels of their foci, even though colors are named as G "red," B "yellow," R "green," and Y "blue." Only K "white" and W "black" would not be detectable in focus selection.

If inversion of W-K to K-W were ascertainable, behavior involving application of their names might provide the clues. Inventories of patterned differences between naming of W versus K are not published; they are compiled only for the Mesoamerican Color Survey (MacLaury 1986). Some are as follows, taking examples from English: (a) the term naming W shows greatest elaboration, for example, "off-white" is standard whereas "off-black" is novel; (b) K is most likely to be renamed, as "blak" once replaced "swarte"; the term naming W applies to more Munsell chips than does the term naming K (e.g., MacLaury 1997b, p. 12, Fig. 1.3). In languages other than English and only rarely, the term naming W also names K (e.g., MacLaury 1991, p. 40, Fig. 9), but the reverse has never been found. The inverted spectrum argument could make its last stand among only W and K in hope of surviving indications such as those.

Palmer does not argue that variation in behavior would mask inversion of W-K, R-G, or Y-B, even though it surely would for many individuals in the messy real world. I maintain his level of abstraction. Otherwise variation in any domain could be said to conceal inversion.

With the tentative exception of W-K, color naming is noticeably asymmetrical, and thus does not lie on the other side of a subjectivity barrier when we consider contrasts as blatant as R-G or Y-B. The incommunicability between your experience and mine of, say, a particular green color chip will be limited to its shade, given that we may differ genetically within bounds allowing nonanomalous color vision. Palmer's arguments concerning functionalist epistemology may prevail on their own. But I do not link them in a privileged way to the naming of color.

## Neurophenomenological constraints and pushing back the subjectivity barrier

Bruce MacLennan

*Computer Science Department, University of Tennessee, Knoxville, TN 37996.* **maclennan@cs.utk.edu** **www.cs.utk.edu/~mclennan**

**Abstract:** In the first part of this commentary I argue that a neurophenomenological analysis of color reveals additional asymmetries that preclude undetectable color transformations, without appealing to weak arguments based on Basic Color Categories (BCCs); that is, I suggest additional factors that must be included in "an empirically accurate model of color experience," and which break the remaining asymmetries. In the second part I discuss the "isomorphism constraint" and the extent to which we may predict the subjective quality of experience from its neurological correlates. Protophenomena are discussed as a way of capturing in a relational structure all of qualitative experience except for the bare fact of subjectivity.

***Via negativa.*** Many of the issues addressed in Palmer's target article can be investigated by a *neurophenomenological* approach, which seeks systematic parallels between the structure of experience, as revealed by phenomenological analysis, and the structure of the nervous system, as investigated by neuroscience (MacLennan 1995; 1996a). The use of phenomenological techniques is especially important if we are to avoid theoretically preconditioned oversimplifications of the phenomena.

Consider first a light/dark (white/black) inversion. The color sphere suggests that this is possible, but a phenomenological analysis argues against it, for the light and dark have different phenomenological structures. As Francis Bacon ("Of Unity in Religion," *Essays*, 1625) said, "All colours will agree in the dark"; that is, all hues merge at the bottom of the color sphere. The sphere similarly shows the hues merging at the top, but this seems to be more an artifact of the theory than a phenomenological reality. At best it is a rare experience, such as one might have staring at the sun or into a very bright light. But this reveals another asymmetry of the light/dark axis, for very bright lights are accompanied by pain, but darkness is not. This experience of pain is an integral part of the phenomenology of vision.

I have little to add to Palmer's treatment of the yellow/blue inversion, except to observe that the "yellow anomaly" (the fact that yellow is inherently lighter than the other colors) is predicted and explained by the fact that the response of the yellow channel $(-S+M+L)$ has the largest overlap with the light (white) channel $(+S+M+L)$ of all the color channels.[1] Here neuroscience complements phenomenology.)

Thus, since ancient times (e.g., Aristotle, *De sensu*, p. 442a), phenomenological analyses of color have recognized the similarity between yellow/blue and light/dark, often making them the ex-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

tremes of a color-series arranged linearly between light and dark. Furthermore, the first two colors in the Berlin and Kay hierarchy are conventionally termed "white" and "black," but are more accurately described as warm-light versus cool-dark, that is, very much like yellow/blue (Kay & McDaniel 1978).

The red/green inversion is more difficult, and so the target article makes a problematic appeal to Basic Color Categories (BCCs); I think there is a better approach, however. By a careful phenomenological analysis of colors, Goethe (1840) was able to identify an important difference between our experiences of red and green. Yellow and blue as extremes can be combined to yield green as a mean (para. 697), which is experienced as similar to both yellow and blue (even though unique-green contains no yellow or blue). Red is not intermediate in this way. Instead, by a process of augmentation (*Steigerung*) of intensity, blue and yellow both approach a very pure red or *Purpur* (para. 699–703), a color "like fine carmine on white porcelain" (para. 792). Blue passes through violet to *Purpur*, and yellow passes through orange (para. 704). Because *Purpur* is not experienced as a simple mixture or union, Goethe classifies it as the third primary color (after yellow and blue). Green, however, is classified as the first secondary color, because it is seen as a mixture of the primaries blue and yellow.

Goethe's analysis is supported by the Berlin and Kay studies (1969), which make red the third color after "black" and "white," and green the fourth. The phenomenological analysis is confirmed by neuroscience, because the green channel (−S+M−L) has a significant overlap with both yellow (−S+M+L) and blue (+S−M−L), but red (+S−M+L) overlaps only yellow significantly (thus it is similar to yellow but not to blue, this may be caused in part by the comparatively small number of S cones.)

In summary, the possibility of a spectral inversion arises from a naive identification of color experience with the linear spectrum. However, progressively more careful phenomenological analyses of color (beginning with Goethe, 1840, and Hering, 1878/1964) have revealed richer structures and imposed additional constraints on possible inversions. I expect this progress to continue. For example, so far as I am aware, there is still no adequate explanation of Benham's disk, the illusion in which colors emerge from a rotating black and white disk. However, a neurophenomenological explanation of this is likely to reveal additional asymmetries in experienced color space. It seems to me that the inverted spectrum is doomed if not already dead.

***Via affirmativa.*** In the second part of my commentary I would like to move from the impossibilities of color inversions to the possibilities of explaining color experience. Thus, we seek to explain the phenomenology of color in terms of the neurophysiology (and also neuroethology!) of normal vision, but one of the tests of such a theory will be our ability to predict experiences associated with abnormal color vision. However, we must consider first what we may expect from such an explanation, and what we may not.

We would, of course, like to be able to imagine the color experiences of people and other animals with color vision significantly different from our own. However, there are good reasons for doubting our ability to do this. In sensory areas of the brain, imaginal layers (with inputs from higher regions) appear to alternate with perceptual layers (with external inputs) and to have parallel structures. Therefore, if the structure of our possible experiences is determined by corresponding neural structure, then our ability to imagine vividly will be likewise limited. That is, neurological structure defines the topological structure of our experiences, both perceived and imagined.

Therefore, it seems unlikely that we can vividly imagine perceptual experiences radically different from our own, a conclusion that seems to be confirmed by everyday experience. What we can hope for are qualitative and quantitative verbal descriptions of the topological structure of alien perceptual systems; we can seek visualizations where they are possible, but we must be prepared to abandon them as we explore perceptual systems progressively more different from our own.

Based on the preceding discussion, we can hypothesize that whatever color channel has the greatest overlap with the light channel (+S+M+L) will be experienced as yellow, or to put it the other way, phenomenal yellow is the experience of the chromic channel with the greatest overlap with light. Indeed, yellow and blue can be considered the chromic correspondents of light and dark ("white" and "black").

In normal color vision, the unimodal channel (−S+M+L / +S−M−L) generates yellow and its opposite, whereas the bimodal channel (−S+M−L / +S−M+L) generates green and its opposite. The unimodal channel generates experiences of yellow because the two adjacent response curves (M and L) combine with a greater overlap than the two nonadjacent ones (S, L) in the bimodal channel; therefore the unimodal channel has the greater overlap with the light-dark channel.

Green and red are the unique hues that are less like light/dark than yellow/blue; this is caused by the lesser overlap of the response curves. The phenomenological structure of green is given by its similarity to both blue and yellow, whereas its opposite, red, is similar to yellow, but not to blue. The green channel (−S+M−L) has a substantial overlap with both yellow and blue, whereas red (+S−M+L) has a substantial overlap with yellow but a much smaller one with blue. Therefore, we may hypothesize that green is the experience resulting from the channel with the greatest overlap with both of the extremes, whereas red results from an overlap with yellow but not blue.

In an abnormal system that had +S+M−L and −−M+L for the unimodal channel, experience of yellows would correspond to spectral blue-green light (+S+M−L), the region of largest overlap with the light channel (+S+M+L), and spectral orange-reds (−S−M+L) would be experienced as blues. Phenomenal greens (−S+M−L) would still correspond approximately with spectral greens, because they have to be similar to both phenomenal yellow and blue, but the opposing phenomenal reds (+S−M+L) would correspond, I think with spectral violets (and nonspectral purples). Such anomalies could be detected by subjects trained in the phenomenological description of their color experience; for example, spectral green light would not be experienced as similar to both yellow and blue.

A more problematic abnormality replaces the bimodal channel with a unimodal channel, so that there are two unimodal channels: −S+M+L (and its opposite) and +S+M−L (and its opposite). The problem is that in the worst case there could be complete symmetry between these channels, so they have equal overlaps with the light channels and equal claim to generate the yellow experience (though for different wavelengths). Topologically each of these channels would appear like yellow in its relation to its opposing blue and in the relation of the yellow-blue pair to light-dark. However, there would be an additional similarity between each blue and the yellow of the other pair. here we may have an example of a visual system too alien to permit visualization of the experience, but the topology is clear. In any case the anomaly would be easily detectable, because there would be only two spectral unique hues, as opposed to the three spectral unique hues of normal color vision.

In section 2.3 of the target article it is noted that the isomorphism criterion arises in mathematics, as well as in behavioral science. Indeed, it arises in any objective science, as traditionally construed. In the end, they are all expressed in terms of external relations among primitive objects. Thus physics says nothing about what electric charge *is*; it limits itself to describing the relations between charged particles and electromagnetic fields.

However, it may be argued (Chalmers 1995; 1996; MacLennan 1995; 1996a) that the traditional approach is inadequate for solving the "hard problem," that is, for fully integrating consciousness into the scientific worldview. This is because an adequate account of subjectivity must deal with certain relations between objects (specifically, those between subjects and the objects of their consciousness) from a different perspective, from the "inside" of the subjects out to their objects. Such relations cannot be entirely external to the objects. Therefore, we must expand our domain to


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

admit that some objects, at least, have two sides (the outside and the inside, so to speak), one of which is accessible only when the observer *is* the object. Thus we are led to some form of dual-access monism.

My own approach to these problems is by way of a theoretical entity, the *protophenomenon* (MacLennan 1995; 1996a, 1999). Protophenomena are elementary units of experience, postulated to correspond to certain brain activity sites (perhaps the somata of neurons). It must be emphasized that protophenomena are very small; the number constituting an individual's consciousness state would be on the order of the number of neurons in the cortex, say thirty billion. Each protophenomenon has a subjective *intensity,* which we may call the *fundamental quale.* The intensities of protophenomena depend on the intensities of other protophenomena (and on extrinsic independent variables) in mathematically definable ways that correspond to the electrochemical connections between neurons (MacLennan 1996b). Protophenomena acquire their qualitative character from these mutual dependencies, which define the structure of possible experiences.

At the end of section 2.2 we read that "the nature of color experiences cannot be uniquely fixed by objective behavioral means, but their structural interrelations can be." I am more optimistic, however. As we come to better understand the neurophenomenology of color we will be able to reduce more of its phenomenological structure to the relations between protophenomena and their neurological parallels. In the end, all that should remain unreduced is the bare ("colorless") fact of subjectivity, represented by protophenomenal intensity.

Furthermore, because protophenomenal intensity is a very simple property, it is possible, at least in principle, to approach many questions concerning experience empirically by means of phenomenologically trained subjects. For example, by controlling the activity at activity sites and having subjects report protophenomenal intensity we may determine whether absolute or relative neural activity corresponds to intensity, which will go toward answering questions such as whether one person is experiencing a "whiter white" than another (cf. sect. 2.3).

In conclusion, neurophenomenology reveals color experience to have a rich structure that precludes color transformations such as suggested by Locke (1690/1987). Further, by showing the parallels between neural structure and phenomenological structure, it allows us to predict the phenomenology of color systems different from our own. If this structuralist approach is correct, then all that is behind the "subjectivity barrier" is the bare fact of subjectivity, represented by protophenomenal intensities; the rich quality of experience will be explicable in terms of protophenomenal dependencies.

NOTE

**1.** The De Valois & De Valois (1993) system is a little different from that used in this commentary. Although they have −S−M+L for the yellow channel, it still has the largest overlap with the light channel, because of the large proportion of L cones; they use the ratios L:M:S = 10:5:1.

## Consciousness – subject to agreement

Neil Law Malcolm

*Clare Hall, University of Cambridge, Cambridge, CB3 9AL, United Kingdom.*
**malcolm@rowland.org** **www.clarehall.cam.ac.uk**

**Abstract:** The claim that isomorphism in perceptual behaviour allows for differences in inner experience holds only if experience is taken to be an entity quite distinct from perceptual behaviour and only accidentally related to it. But this is not so. The two are internally related; experience as conceptualised being inherent to perception as a species of normative behaviour.

My reaction to reading the target article is that Palmer has reached a correct conclusion – how a colour looks to someone, its subjective character, is immune to strict behaviourist or biological explanation – but for the wrong reason (that experience lies behind an objectively impenetrable "subjectivity barrier," sect. 2.1). Because of this, Palmer is led to the further, erroneous conclusion that one can never know another's colour experiences from that person's behaviour, even in principle (sect. 4, para. 2). His argument for this mistakes the nature of colour experience, its links with perceptual behaviour, the role of language in identification, and the appropriate mode of investigation into consciousness of colours.

***1. Colours are not logically private objects of experience, but are objective properties of things in the world around us.*** Palmer's problem of how we can know what colours another experiences is a relic of his conceiving of subjective experiences as "internal . . . private events" (sect. 1.1, para. 5), which are entirely independent of perception and conceptualisation and only contingently, that is, causally, connected to behaviour. Now the matter of whether I can be sure that you and I are having the same experience on looking at this page is like the matter of whether I can be sure that my current judgment of white is the same as my previous one: there must be criteria for seeing white as the *same again.* But if colour experience is a private entity to which only its possessor has cognitive access, the criteria for reidentification can only be arbitrary and, as Wittgenstein (1953) famously pointed out, "whatever is going to seem right to me is right. And that only means that here we can't talk about 'right'." Were I a speaker of a *private* language, no possibility would exist of my checking, verifying, or being wrong about colours because there would be no difference between my being right and being wrong, except that it seemed a certain way to me. Only if I am a speaker of a *public* language can I attach meaning to the concept of colour. This allows the logical possibility of others, or oneself, checking, verifying, or being wrong about colours, which means it makes sense to say my judgments are right or not. What we see when we see something white is not an inner experience, but an objective, irreducible property of that public thing. The easy equation of colour with experience stems from the refusal to countenance the objectivity of colours (Malcolm 1999).

***2. There is an internal relation, mediated by colour concepts, between experience and behaviour.*** The relation between subjective experience and perceptual behaviour is not an entirely accidental one between distinct entities. Nor is it one of identity. Rather, it is internal, neither experience nor behaviour being properly understood apart from the relationship in which they stand. Because colours, properly construed, can only be seen as conceptualised, the link between the two is mediated by the concepts of the colours that we share with others. This means that other people's colour experiences are not irrevocably hidden from us but are manifest in and through their behaviour (while still being more than just behaviour). Remember, the terms of the subjectivity debate are not over whether you and I can *have* the same experience in the sense of suffering the one experience; clearly we cannot. They are about whether I can *know,* from my third-personal vantage, what you are experiencing from your first-personal vantage, specifically, the character of your colour experience. And the answer is yes, provided the relevant sense of "colour" in the expression "colour experience" is taken to be "colour as conceived." If perceptual behaviour is the natural expression of experience as conceptualised, then, although I cannot physically have your experience, I can logically know what your experience is, given its behavioural expression in terms meaningful to me.

***3. Colour language is the means of conceptualising experience in agreed ways.*** Colour language, *pace* Palmer (sect. 2.2, para. 3), does not provide a set of labels for what we privately experience and recognise, but functions as the means of conceptualising what its speakers experience. Concepts are capacities exercised in acts of judgment, and acts of colour judgment presuppose the possession of colour concepts. Without concepts we can neither think about the colours nor identify them, only differentiate coloured things in ways permitted by the workings of



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

our visual system. A sufficient condition for our having concepts of the colours is that we have mastered the intelligent uses of colour-words in some language. If colour experience is naturally expressed in and through what one says and does then, so long as your judgments invoke the use of agreed concepts, I can know what colours you are seeing. Importantly, the basis for this knowledge is public and available; it involves intersubjective agreement about the rules for the use of colour language and how these rules are to be applied and it implicates samples of colours. Our mastery of a shared vocabulary and grammar of colour therefore enables us to see the colours – *have the same experiences* – that others do.

***4. The nature of subjective experience – consciousness – or colours is a matter for conceptual investigation.*** Attempting to "get a scientific handle on [the] philosophical problem of whether transformed color experiences could be detected in publicly observable behavior" (sect. 1, para. 4), Palmer reaches the solipsistic conclusion that "there is no way to specify uniquely the qualities of particular experiences except by reference to one's own" (sect. 4, para. 9). But the reason that colour experience is recalcitrant to scientific explanation is that there are no conceptual connections linking neural states or purely physical behaviour to consciousness of colours. Science is wholly unsuited to the explanation of conceptualised experience, which is what is at issue. The inaccessibility Palmer adduces is strictly an empirical concern. Subjective experience – seeing as from a first-personal point of view – is, however, accessible to *conceptual* inquiry which can tell all there is to know. How in practice do we realise what state of mind someone else is in? We do so not by inference or analogy, but in what that person says and does under the circumstances, where what counts is not "hard" physical behaviour, but meaningful, communicative *human* behaviour enacted for a purpose in social and visual context. We take another's behaviour *from the start* as expressive of mind. Subjective experience or consciousness of colours is inherent to perception as a species of normative behaviour (Malcolm, in preparation).

## Beyond intrinsicness and dazzling blacks

Erik Myin

*Department of Philosophy and Artificial Intelligence, Free University Brussels, B1050 Brussels, Belgium.* **emyin@vub.ac.be**
**homepages.vub.ac.be/~emyin/**

**Abstract:** The concept of intrinsic aspects of color experience is flawed and functionalism remains plausible. The idea of spectrum inversion cannot survive in the context of a realistic conception of color.

Palmer's target article is surely one of the most scientifically detailed and knowledgeable treatments of spectrum inversion ever. Unfortunately, it is built on a very shaky philosophical foundation, the notion of the "intrinsic." In the article's ontology, there are two kinds of properties of mental states, intrinsic properties and relational properties. The whole point of the article is that these aspects of experience are mutually exclusive: The intrinsic is nonrelational and the relational is nonintrinsic.

It is difficult to make sense of the notion of intrinsic aspects of color experience. Take, for example, a sensation of orange. According to the logic of the target article, the relational aspects of this color sensation would involve its similarity relations to other color sensations (such as that it is more similar to red than to blue) and facts concerning its compositional structure (such as that it is composed of the Hering primaries red and yellow). But besides these relational aspects, and completely independent of them, there would be a still further "intrinsic" aspect. What experiential content could such an aspect carry? We can only approach this question negatively, by looking at what is left over after subtracting all relational aspects. This means the "intrinsic" content would *not* be experienced as "composed" and it would *not* be experienced as being related to other colors. It is hard to see, however, how such a content still specifies anything that would be "orange-like."

A more general worry concerns how intrinsic aspects could figure in experience at all. For is it not the hallmark of experience that it is subjective, thus experienced as related to the experiencing subject? Subjectivity, being "for a subject" seems to be preeminently relational (cf. Church 1998). If the category of the "intrinsic" specifies anything at all, it surely cannot be something that matters for consciousness! Of course one can launch the category of the intrinsically subjective, but this seems a merely verbal and desperate move. The fact that one can give a name to a paradoxical category does not make it any more viable.

A favorite strategy of believers in the "intrinsic" is to dismiss any critique of it as "eliminativism," which then gets further portrayed as the denial of the existence of consciousness, despite the fact that critics of the intrinsic are the first to offer – relational! – theories of consciousness. There is little room for developing such a theory for color experience here, but it should certainly include the following aspects:

1. First-order representations that code for color. These can stand in various relations to each other, such as composition or incompatibility.
2. Higher order functional entities that are sensitive to these relations (these could be further representations or first-order representation using mechanisms).
3. Representations of body and self in terms of which incoming representations can be made sense of in "subject-centered" coordinates (their application implies conscious content is "for the subject" and can guide its actions. Representations seem necessary even for the body because of its distance to the brain) (Damasio 1994).

In the target article only representations of type (1) are talked about, as if they were sufficient for consciousness. This quickly leads to such pseudoproblems as whether a camera connected to a color labeling computer has color sensations.[1] The sufficiency of type (1) representations is refuted empirically by the fact that spectral sensitivity curves measured for color blindsight have their normal "opponent" form (Stoerig & Cowey 1992).

In a realistic functionalist theory of color experience, color consciousness would arise out of the interactions of various functional components (such as 1, 2, and 3 mentioned above) and no component by itself would suffice for consciousness. The color room fantasy only works against – to my mind nonexistent – functionalist theories where the denial of this last assertion is embraced.

But what about spectrum inversion? Should we not address this problem directly, instead of attacking it indirectly by undermining intrinsicness? By dismissing the intrinsic, a defender of a functionalist theory would indeed have to construe colors in relational terms. In the end the identity of a color would be constituted by its collective set of relations (Harden 1988). However, the set of relations should not be restricted as in the target article to the three dimensions of psychophysical color space (which are to a certain degree artificial, because they describe color experience only under very narrowly delimited conditions), but could include relations with anything one can think of, such as: sensations in other modalities, emotions, colored objects, and physical properties of light.

Questions of symmetry become immensely more complicated within this holistic picture. Consider, for example, the property of being dazzling, and the fact that there can be no dazzling black. This is a result of optical facts concerning the interaction of light and objects, essentially the fact that black objects are those that absorb all incoming light and that being dazzling implies the reflection of light from objects. The "nondazzlingness" of black is thus both experiential (it determines the nature of our experience of black) and physical (it is determined by physics). The impossibility of a dazzling black seems to make a reversal of the lightness axis – treated as unproblematic in the target article – impossible,


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

for such a reversal would create precisely the category of a dazzling black. Even if he considered the physical impossibility involved as irrelevant, the defender of spectrum inversion would have to grant detectability here!

Broadening our concept of color to include more of its relations roots color more firmly in both the mind and in the world and brings to the fore the outlandish nature of fantasies such as Locke's spectrum inversion.

ACKNOWLEDGMENTS

Thanks to the Flemish Community (grant CAW 96/29b) and the Free University of Brussels (VUB-project GOA 2) for financial support.

NOTES

**1.** At the Artificial Intelligence Lab of the Vrije Unversiteit Brussel, Luc Steels and coworkers – including me – are performing experiments with color camera-connected programs that very closely fit the description offered in the target article. Given the relative simplicity of this setup, our research goal is to investigate artificial color categorization and certainly not artificial color consciousness.

## Normal, pseudonormal, and color-blind vision: Cases of justified phenomenal belief

Martine Nida-Rümelin

*Department of Philosophy, University of Fribourg, CH-1700 Fribourg, Switzerland.* **martine.nida-ruemelin@unifr.ch**

**Abstract:** Palmer's "isomorphism constraint" may be interpreted as a claim about (1) what can be *known with certainty,* or (2) what can be *detected,* or (3) what *beliefs* can be *justified* on the basis of a certain kind of scientific knowledge. I argue that his claim is valid if interpreted in one of the first two ways, but invalid if interpreted in the third way.

Palmer's claims about the difference in our epistemical status with respect to phenomenal structure on the one hand and phenomenal quality on the other may be reformulated in three different ways: On the basis of behavioural, functional, and biological data:

(T1) We can *know with certainty* that another person's experience has certain structural properties, but *not know with certainty* that it has a specific intrinsic quality.

(T2) We can *detect* structural properties of another person's experience, but not its intrinsic quality.

(T3) We can form *justified beliefs* about structural properties of another person's experience but not about its intrinsic quality.

***Comment on T1.*** It is not obvious on what reading of "knowing with certainty" both parts of the claim could turn out to be valid. If "knowing with certainty" requires that every possible alternative be logically or conceptually excluded, then the notion is probably too narrow for the first part of the claim to be valid. If it requires only that the hypothesis at issue be "beyond a reasonable doubt," then the second part of T1 may be wrong (see my discussion of T3 below). Despite these difficulties one should, I think, agree with Palmer that there is a kind (or degree) of certainty achievable in the first case (phenomenal structure) that cannot be achieved in the second (phenomenal quality).

***Comment on T2.*** Those who do not know a given intrinsic quality (e.g., the specific quality called "red") from their own experience are – in a certain sense – unable even to *consider* the question of whether another person's experience is an experience of this specific kind (e.g., red). What cannot be considered by a given subject, cannot be believed or known either. This quite radical epistemic inaccessibility of facts about an intrinsic quality of experience for a subject not acquainted with the quality at issue cannot be overcome by acquiring behavioural, functional, and biological knowledge.[1] Knowledge about phenomenal structure, by contrast, can be acquired without acquaintance with the specific type of experience at issue. On a natural understanding of "detecting," the claim that intrinsic quality can be "detected" on the basis of "objective" data assumes that one can gain phenomenal knowledge about the intrinsic quality of another person's experience even if not acquainted with that quality oneself. If we read "detect" in thesis T2 in this way, then T2 should be accepted.

***Comment on T3.*** Palmer seems to accept implicitly a quite strict necessary condition for justified phenomenal belief. He writes: "if there are any potentially relevant differences between our brains that might produce experiential differences, it is unjustified to assume equivalence of color experiences" (sect. 3.4, para. 5). This quotation still leaves open the question of whether he thinks my belief that you have an experience of green is justified only if:

(1) I have no reason to assume that there are potentially relevant differences between your brain on the occasion at issue and my brain when I have an experience of red, or

(2) I have reason to assume that there are no such differences.

In Palmer's immediately following lines it seems quite clear, however, that he tends to choose the second alternative and that he requires much for its fulfillment. He seems to think that the phenomenal belief at issue is unjustified unless I have already established (a) what differences are potentially relevant, and (b) that there are no such differences on the relevant occasions between you and me. I would instead propose the following sufficient condition for a scientifically based justified phenomenal belief: My phenomenal belief that you have (e.g.) a sensation of green is justified if:

(3) I have reason to believe that there is a specific type of physiological process R that is in my case responsible for sensations of green, and

(4) I have reason to believe that a process of this type R occurs in your brain on the occasion at issue, and

(5) I have no reason to believe that there are potentially relevant differences between your brain when a process of type R occurs and my brain when a process of type R occurs.

Note that I can fulfill (5) even if no empirical results about potentially relevant factors like those mentioned by Palmer in section 3.5 are at hand. According to the proposed sufficient condition, there are plenty of cases of scientifically justified phenomenal beliefs: (a) the belief of a normally sighted person that red, green, yellow, and blue are the basic hues experienced by all normally sighted people, (b) the belief that pseudonormal people are red-green inverted[2], or (c) the belief that completely red-green blind people have yellow and blue as their only basic hues.

NOTES

**1.** Nagel (1974) and Jackson (1982) are well-known advocates of this view (e.g., compare Jackson. A precise formulation requires, in my opinion, the distinction between two kinds of belief (non-phenomenal and phenomenal belief) about qualities of experience introduced and discussed at length in Nida-Rümelin (1998). Phenomenal belief about a quality of experience requires that the epistemic subject be acquainted with the quality at issue. I use the notion of phenomenal belief in the following, although there is no room to introduce the distinction here.

**2.** Pseudonormal people have their R-receptors filled with the photopigment normally contained in G-cones and their G-receptors filled with the photopigment normally contained in R-cones. Their existence is predicted by Piantanida's theory about the inheritance of red-green blindness (see Piantanida 1974 and Boynton 1979). It follows from central assumptions about the physiological basis of color vision that these people have normal discriminative capacities but have their red- and green- sensations reversed. For further information about the hypothesis of pseudonormal vision and a discussion of related philosophical questions see Nida-Rümelin (1996; 1999).



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

## Finding a place for experience in the physical-relational structure of the brain

Gerard O'Brien and Jonathan Opie

*Department of Philosophy, University of Adelaide, South Australia 5005, Australia.* **gerard.obrien@adelaide.edu.au jopie@arts.adelaide.edu.au arts.adelaide.edu.au/Philosophy/{gobrien.htm; jopie.htm}**

**Abstract:** In restricting his analysis to the causal relations of functionalism, on the one hand, and the neurophysiological realizers of biology, on the other, Palmer has overlooked an alternative conception of the relationship between color experience and the brain – one that liberalizes the relation between mental phenomena and their physical implementation, without generating functionalism's counter-intuitive consequences. In this commentary we rely on Palmer's earlier work (especially from 1978) to tease out this alternative.

What can natural science hope to tell us about the qualitative character of experience? Palmer is pessimistic: Not only is behavioural science unable to explain the qualities of individual experiences, it cannot even distinguish those systems that have experiences from those that merely simulate them. Biological science is not much better off, he claims, because it cannot even tell us when two subjects are having the same experiences (save for those exceedingly rare cases where neurophysiological identity obtains), let alone what these experiences are like.

Given this pessimistic assessment, it is somewhat ironic that Palmer himself has developed some conceptual tools that, when applied to experience, may extricate natural science from this predicament. In a paper setting out some fundamental features of representation, Palmer observes that there are two different means by which a set of (representing) objects can preserve the relational structure of another (represented) set: "Representation is . . . *intrinsic* whenever a representing relation has the same inherent constraints as its represented relation . . . [On the other hand] representation is . . . *extrinsic* whenever the inherent structure of a representing relation is totally arbitrary and that of its represented relation is not" (1978, p. 271, the emphasis is ours).

Let us apply this distinction to the brain's representation of color. Palmer tells us that "the entire structure of color space . . . is determined by relations among colors, particularly relations of composition and similarity" (sect. 2.2, para. 2). One way the brain might preserve this relational structure, and in so doing generate our color experiences, is via a corresponding set of *causal* relations among its representational vehicles. In other words, it might be that color experiences are nothing more than causal roles, as functionalists suggest. This would render the representation of color *extrinsic,* because with enough ingenuity the appropriate set of causal relations can be imposed on just about any set of representing objects, regardless of their inherent structure. It is precisely because it involves extrinsic representation that functionalism renders physical implementation irrelevant to mentality, something Palmer is keen to emphasize: "functionalism treats mental phenomena as independent of their physical realizations: Any set of physical events will do, provided they have the right causal relational structure" (sect. 4, para. 6). And it is this that he thinks undermines the capacity of functionalism to explain color experience (sect. 4, para. 6–8).

Another way the brain might preserve the relational structure of color space is by employing corresponding *physical* relations among its representational vehicles. This would render the representation of color in the brain *intrinsic,* because such physical relations obtain only in virtue of the inherent physical structure of the representing objects. Palmer canvasses this option in section 3, when he examines the biological conception of mental phenomena: the identification of conscious experiences with neurophysiological states. The difficulty here is the fact that "most people's brains differ from each other in a multitude of ways," making it look extremely unlikely that we will be able to discover a "principled physical correspondence" (sect. 3.4, para. 3) on which to ground a theory of color experience.

However, there is a way of understanding color experience in terms of intrinsic representation that does not seek to identify mental phenomena with neurophysiological states. Again, the basis of this idea is to be found in Palmer (1978). He notes that two systems can implement the same set of physical relations without being physically identical (pp. 296–97). This suggests that we hazard an identification of conscious experiences with what might be termed structural roles, which are defined in terms of the physical relations among a set of representing objects. Crucially, physically distinct objects can play the same structural role within different representational systems, so long as they bear the same physical relations to other members of their respective systems. Given that this approach relies on the structural roles played by the brain's representational vehicles, rather than their particular physical properties, we propose to call it *structuralism.*

The distinction between structuralism and the biological conception of mental phenomena can be illustrated by reference to parallel distributed processing (PDP) systems. It is well known that two implementations of, say, NETtalk, can vary considerably at the level of weights and connections (the "biological" level), yet still perform the same mapping of text onto phonemes (Churchland 1998; Sejnowski & Rosenberg 1987). Some theorists see this as a vindication of functionalism. Distinct NETtalks appear to be merely input-output equivalent. But if one digs deeper, it emerges that this similarity in causal profile is grounded in an underlying structural similarity: Cluster analysis reveals that different implementations of NETtalk partition activation space *in the same way.* This indicates that the set of physical relations among hidden layer activation patterns (as codified by distances in activation space) is preserved across distinct implementations of NETtalk, even if they vary at the level of weights and activation patterns.

Structuralism is therefore best understood as occupying a theoretical position midway between functionalism and the biological conception. Because two representational systems can share the same set of physical relations without being physically identical, structuralism, like functionalism, liberalizes the relation between mental phenomena and their physical implementation in the brain. Unlike functionalism, but in common with the biological conception, structuralism makes physical implementation relevant to mentality: It takes mental phenomena to be constituted by the physical relations their realizers bear to one another, not their causal relations. Thus, structuralism has all the virtues of these traditional rival theories of mind, without possessing their vices.

When it comes to the explanation of color experience, it is important to note first that structuralism is capable of being borne out (in principle) by within-subject experiments involving physiological interventions. If we discover that the color experiences of an individual are not affected by structure-preserving changes in brain circuitry, but *are* affected by alterations to inherent (physical-relational) structure, we can reasonably conclude that color experiences are identical to structural roles. In this circumstance similarity at the structural level trumps functional similarity, and color-zombies can be safely identified by their lack of appropriate representational structures.

One might object that insofar as structuralism offers a relational analysis of experience, just like functionalism it fails to address the subisomorphic features of color experience. In particular, what is to prevent two people with color representation systems possessing the same physical-relational structure from having qualitatively different color experiences? However, there is no more reason for thinking that two such people could have different color experiences than there is for thinking that neurophysiological clones could have different color experiences – the latter being something that Palmer concedes as being unparsimonious (see sect. 3.4, para. 2). The transformed color thought experiment really only starts pumping out its famous intuition when the relational structure at issue is causal.

The deepest source of Palmer's pessimism revolves around the incapacity of natural science to explicate the qualitative character



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

of individual color experiences. Can the structuralist conception of mentality give us any purchase on this last bastion of subjectivity? Perhaps Palmer is right that objective science reaches an ultimate barrier at this point. But structuralism at least holds up the prospect that the equivalence class of subjects with identical color experiences is far larger than the biological conception allows. A person can thus understand (in the relevant first-person sense) the color-experiences of another person if their respective color systems are structurally identical, even if there are quite significant physical differences between their color representations.

## One basic or two? A rhapsody in blue

Galina V. Paramei

*Institut für Arbeitsphysiologie an der Universität Dortmund, 44139 Dortmund, Germany (on leave from the Institute of Psychology at the Russian Academy of Science, Moscow, Russia).* **paramei@arb-phys.uni-dortmund.de**
**www.ifado.de/projekt-06/**

**Abstract:** The controversial status of *goluboi* as a basic color term is discussed. Fuzzy logic alone cannot reliably attribute basic status to *goluboi.* Recent linguistic studies support a single basic blue category. Psychophysical data on color-space distances and color naming are currently ambiguous in this regard.

Among derived basic color categories (BCCs), Palmer lists Russian *goluboi,* "light blue" (sect. 1.4). The term is not listed among basic color terms (BCTs) identified by Berlin and Kay (1969), and its basicness is considered disputable. In the context of an inverted spectrum, if *goluboi* were deemed a BCT, four more symmetries in color experience would remain unbroken under reflectional transformations – those of *brown* and *goluboi* under B-Y/Bk-Wh and under R-G/B-Y/Bk-Wh inversions (see Table 1 of the target article).

The inclusion of *goluboi* among derived BCTs would be reasonable following the logic of fuzzy-set membership functions, to which Palmer adheres. Derived color categories result from the intersection of two primary color categories, specifically, *goluboi* of *white* and *blue.* The conjoining of some primary categories, however, also gives rise to nonbasic terms and as such this operation cannot automatically warrant basicness.

Palmer's assumption that the basic status of derived terms may result from the wider perceptual gap between particular unique hues can be tested psychophysically, assuming that referents are defined as samples of the most representative color of those hues (cf. Saunders & Van Brakel 1997, p. 168). Indeed, a "color circle" reconstructed from large color differences between Munsell samples (Indow 1988) demonstrates that the *orange* (5YR) is located at the midpoint between widely separated unique red and unique yellow, whereas *purple* commensurably bisects a great distance between unique red and unique blue; nonbasic *cyan* (5BG) lies at the intersection of closely-located *blue* (5B) and *green* (5G). Analogously, the basicness of *gray* is explained by its location between polarized samples of *white* and *black.* Psychophysical evidence for the basicness of *pink* and *brown* can be adduced from a study by Fenton (1997), who found *red* to be most distant from *white* (as well as from the other primary colors), and *yellow* farthest from *black.* Although Fenton found *black* to be closer to *blue* than to *red* or *yellow,* this finding provides insufficient evidence that the separation between *blue* and *white* is large enough to foster emergence of a BCC at their intersection. Thus, on psychophysical grounds the feasibility of *goluboi* basicness is uncertain.

In the linguistic domain, Corbett and Morgan (1988) and Davies et al. (1991) maintained that two Russian terms, *sinii,* "dark blue," and *goluboi,* "light blue," meet linguistic criteria for basicness in their frequencies of occurrence and derivational elaborations. However, on testing perceptual-cognitive relationships, the authors recanted, concluding that the two terms in question may not refer to completely separate perceptual categories, and that "universal 'blue' . . . may remain as a unitary perceptual basic category" (Laws et al. 1995, p. 88).

The existence of two "blue" basic color categories is disputed by MacLaury (1997b), who argues that the initial conclusions of Corbett, Morgan, and Davies rest only on *salience* measures of the terms. But for the term to be basic, along with being salient, it also needs to be *general,* that is, its meaning should not be subsumable under the meaning of another term. If *goluboi* were basic, its core meaning would stand apart from that of *sinii.* This appears not to be the case: Taylor et al. (1997) used the three-part ethnographic method of assessing overlap among color-term ranges; they posed relations of coextension, inclusion, and polarized inclusion among *sinii* and *goluboi* for different speakers in their sample, but with *sinii* consistently dominant and basic and *goluboi* recessive and nonbasic; the latter became salient without becoming independent.

Further data come from color naming of monochromatic lights obtained from native Russian speakers. Paramei and Cavonius (1997) collected data from Muscovites, using *sinii* for "blue" and adding "white" to the available basic hue terms. The *sinii*-naming function at low luminance (2 cd/m$^2$) was in strong agreement with the "blue"-naming function obtained by Gordon et al. (1994) at comparable illuminance (20 td). But at higher luminances (20 and 200 cd/m$^2$), values of *sinii* function were lower than those of the "blue" function at comparable illuminances in previous studies in which the "chromatic" format was used (Boynton et al. 1964; Uchikawa & Ikeda 1987), being partly substituted by "white"-naming. We are aware that one has to distinguish between the hue components discernible in a sample and the color category to which this sample is assigned, for discernment of hue does not exclude coverage of a certain subset of lights by a BCT (as, for example, red and yellow components are discerned in orangist samples that pertain to a basic *orange* category). Hence, though in our data short-wavelength lights were fully specified by *sinii* and "white" components, they cannot provide a decisive argument for or against a second independent "blue" category.

The issue was investigated in another color-naming study with Russians (presumably in the USA), who named monochromatic lights in Russian or in English (Abramov et al. 1997). When naming in Russian, subjects were allowed to use the basic hue components, with the terms for "blue" allowed in different sessions as *sinii* only, *goluboi* only, or both *sinii* and *goluboi.* When *sinii* was the permitted term, the naming function agreed closely with the "blue"-naming function, implying that *sinii* is the equivalent of "blue" and describes that unique sensation. But with both terms permitted, *sinii* was found to correspond to "blue" + "red" and *goluboi* to "blue" + "green," which the authors regard as supporting the conclusion that both are BCTs. It is noteworthy that the obtained components of the *goluboi* category are at odds with the *blue* and *white* to be inferred from the fuzzy-logic model, whereas these components better match nonbasic *cyan* at the intersection of *blue* and *green.* The discrepancy is explained by the set of terms used, but it might also be caused by low photopic light of exposed stimuli, because the illuminance was only 25 td ( J. Gordon, personal communication). To clarify whether the *sinii*- and *goluboi*-functions – as expressed through the English-named hue components – name two formidably distinct "blue" sensations, further study is needed: The "white"-term must be permitted as a naming option; luminances should be varied to cover the whole photopic range, including high levels that are expected to induce the "white" component.

Although fuzzy logic might portray *goluboi* as a basic color term, it is questionable whether this model verily represents operations of basic-category formation. The most recent linguistic studies provide evidence for a single "blue" basic category, whereas presently available results of psychophysical studies have confirmed neither this conclusion nor its antithesis of two basic blue categories.



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

## Phenomenal experience and science: Separated by a "brick wall"?

Michael Pauen

*Institut für Philosophie, Phillips-Universität, Marburg, 35032 Marburg, Germany.* **pauen@mailer.uni-marburg.de**
**staff-www.uni-marburg.de/~pauen/homepage.htm**

**Abstract:** Palmer's principled distinction between first-person experience and scientific access is called into question. First, complete color transformations of experience *and* memory may be undetectable even from the first-person perspective. Second, transformations of (say) pain experiences seem to be intrinsically connected to certain effects, thus giving science access to these experiences, in principle. Evidence from pain research and emotional psychology indicates that further progress can be made.

Palmer's central argument is based on a fundamental distinction: Although we have direct access to phenomenal experience like color sensations from the first-person perspective, science is restricted to the *relations* between these sensations, thus being separated from experience itself by a "brick wall." I will argue that both claims can be challenged, thus leveling the difference between science and first-person experience and leaving a better chance for explaining these experiences than Palmer wants to concede.

First, Palmer assumes that we can detect color transformations from the first-person perspective. This is not beyond question: Think about a wholesale transformation of color experience *and* memory. Chances are that I will not detect such a transformation because colored objects like tomatoes will appear the way they have always appeared to me–at least, that is what my memory would tell me, and, given the isomorphism constraint, there is nothing and no one to correct it. Even worse, if it occurs to me that the spectrum has changed, it will be unclear whether this is a "real" transformation in my actual experience or just a change in my memory. Thus, my own judgment would be relational, too, because it depends on the relation between memory and experience, rather than on the intrinsic properties of experience.

Second, the actual consequences of the isomorphism constraint depend on whether there is a clear-cut distinction between intrinsic and relational properties of experience. Palmer's central claim is based on such a fundamental distinction, thus completely different experiences can act as "role fillers" for a given functional description, leaving science no chance to detect the difference.

Palmer himself seems to concede certain connections between experience and relational properties of the color space in section 2.2, thus leaving room for the idea that a (presently unavailable) *appropriate* relational description might be able to fix experience itself rather than an extrinsic property of it. Second, an even stronger objection to Palmer's distinction arises from observations on pain. Palmer's claim concerning the difference between experience in general and functional properties would also require that a "feeling transformation" from pain to an opposite experience, say, joy, could occur without an effect on pain behavior. Now, take the case of a newborn baby with such a "feeling transformation": I think the intuition is that the baby will start to cry as soon as it has a pain experience – even if the pain experience is caused by its mother's hugging. This idea is supported by empirical evidence: The prospect for children who do not experience pain is very bad indeed (Pöppel 1995, p. 237).

It would seem then, that at least in cases like this, certain relational properties are constitutive of the *experience itself*: According to our intuitions, experience will *not* change without affecting the relational properties. Current research in neurophysiology indicates that relational analyses can be pushed even further and can connect distinct aspects of pain experience to certain functional properties: Whereas the sensory aspects of pain states are connected to the discrimination of the nociceptive stimuli, the affective aspects are related to motivation (e.g., avoidance behavior, Cross 1994; Rainville et al. 1997). Similar results are provided by recent psychological research on emotions: bodily perception and action tendencies seem to be a part of emotional experience (Damasio 1994; Frijda 1993).

The assumption that phenomenal experiences have functional properties as their constituents is supported by some considerations concerning the causal connection between mental states and brain states. Take the case of memory. I assume that episodes of subjective experience are causally connected to the memory traces of these episodes in the brain. Moreover, the causal connections have to be specific: Phenomenal states of a certain type should reliably cause memory states of a corresponding type; otherwise, I might remember a pain state as an experience of fear or vice versa (Pauen 1999). Palmer himself seems to assume that different experiences differ as far as the associated brain events are concerned, but these differences may be infinitely small. But regardless of how big or small these differences may be, they have to serve as an explanation of the functional differences that Palmer himself acknowledges, thus constituting another connection between experience and functional properties.

Palmer objects that we cannot infer the actual phenomenal experience of a given subject from these variations as long as we do not belong to the same equivalence-class. As far as emotions are concerned, this objection is susceptible to the above argument against transformations: If the experience differs, the relevant functional properties should differ, too. But what about colors? Following the first part of the target article, only three highly specific color transformations are possible. Now, because variations of biological systems like the brain tend to follow a normal distribution, we would expect a lot of small divergences from the average values and only a few big divergences, as they are required by the possible transformations. So, if there *are* color transformations at all, we should find different sorts of transformations, not only those three that might pass through undetected by functional criteria. Conversely, if we do not find detectable divergences then it is highly unlikely that there are divergences at all.

So the conclusion would be that the gap between first- and third-person accounts of phenomenal experience is far smaller than Palmer expects it to be. First, even in the first-person perspective, changes in color experience can pass through unnoticed. Second, the dissociation between experience and functional properties is not as strict as Palmer has it. Thus, it may turn out that the "brick wall" between science and experience is not as impenetrable as Palmer thinks it is.

## An externalist approach to understanding color experience

Peter W. Ross

*Department of Philosophy, California State Polytechnic University, Pomona, Pomona, CA 91711.* **pross@pomona.edu**

**Abstract:** Palmer demarcates the bounds of our understanding of color experience by symmetries in the color space. He claims that if there are symmetries, there can be functionally undetectable color transformations. However, even if there are symmetries, Palmer's support for the possibility of undetectable transformations assumes phenomenal internalism. Alternatively, phenomenal externalism eliminates Palmer's limit on our understanding of color experience.

Palmer argues that there are limits on a scientific understanding of color experience on the basis of possible symmetries in the psychological color space. He points out in his Figure 3A that it is possible that the red-green poles of the color space are reflectionally symmetrical about the yellow-blue axis. Consequently, on the basis of the possibility of symmetries, Palmer purports to establish the possibility of isomorphic color transformations. Because a functionalist account of color experience drawn solely in terms of qualitative relations cannot distinguish between isomorphic color


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

transformations, the possibility of such transformations indicates that this account cannot provide an adequate objective characterization of particular color experiences (sect. 4, para. 2).

Palmer's objection to functionalism does not rest merely on the possibility of *symmetries* in the color space. Rather, Palmer's objection is that functionalism cannot distinguish among certain possible color *transformations.* However, Palmer's move from the possibility of symmetries to that of functionally (and therefore behaviorally) indistinguishable color transformations relies on the controversial assumption that color experience supervenes on neurophysiological properties of perceivers (sect. 3.4, para. 2), a view I will call *phenomenal internalism.*

For example, in section 1.3, paragraphs 4–5, having put forward the possibility that the red-green poles of the color space are reflectionally symmetrical, Palmer appeals to an argument offered by Martine Nida-Rümelin (1996) in support of the possibility of functionally indistinguishable red-green color transformation. Nida-Rümelin points out that an explanation of red-green color blindness indicates that there can be a switch in the photopigments normally contained in the M and L cones, a condition called *pseudonormal vision* (for further discussion of this condition, see Boynton 1979, pp. 355–58). She then contends that pseudonormal vision would be sufficient for functionally indistinguishable red-green transformation, and that because instances of pseudonormal vision are probable, so are cases of such transformation.

However, this reasoning assumes internalism. Clearly, if internalism is correct, and color experience supervenes on neurophysiological properties, a switch in photopigments may be sufficient for red-green color transformation. Such a transformation would elude a functionalist account of color experience solely in terms of qualitative relations.

At least some functionalist accounts (see, for example, Dretske 1995, chap. 5) take an alternative view, called *phenomenal externalism,* to the effect that color experiences supervene on relations between neurophysiological properties of perceivers and colors, where colors are identified with physical properties of physical objects. On this alternative view, the property red is a physical property of objects, and a switch in neurophysiological properties *would not* be sufficient for red-green transformation. Rather, this physical property of being red would be encoded by different neurophysiological properties between those with normal and those with pseudonormal vision. Thus according to externalism, even if there are symmetries in color space, Palmer's support for the possibility of color transformations fails.

Furthermore, functionalists who are also externalists offer an account of color experience that avoids Palmer's objections. His color room thought experiment (sect. 2.5) poses the question: How can computational processes be sufficient for color experience? Palmer insists that even if the computational processes of our visual systems can be mimicked in such a scenario, the qualitative aspect of color experience is surely left unaccounted for. He thus concludes that functionalism fails.

However, because he assumes that functionalism must account for this qualitative aspect *solely* in terms of *internal* properties of perceivers, Palmer ignores the possibility that color experiences are *also* determined by their relations to properties in the world. Indeed, externalist functionalism points to just this possibility, accounting for the qualitative aspect of color experience in terms of its relations to physical properties of physical objects.

Moreover, Palmer's claim that the qualitative aspect of color experience is "purely subjective" and scientifically intractable (sect. 2.1, para. 1) relies in internalism. Certain aspects of color experience – such as whether one has normal or pseudonormal vision – are functionally indistinguishable. Neurophysiological investigation is necessary to determine whether one's photopigments are switched.

However, Palmer's claim that the qualitative aspect of color experience is subjective simply assumes that color experiences supervene on neurophysiological properties. On this assumption, the possibility that the red-green poles of the color space are reflectionally symmetrical, along with the possibility of pseudonormal vision, are all that is needed to prove the possibility of functionally indistinguishable color transformations. With the possibility of such color transformations purportedly established, Palmer concludes that we cannot have knowledge of the color experiences of others who are neurophysiologically different from us in relevant ways. He claims that the only way that we *could* have such knowledge is to have first-person access to their internal states. Lacking such access, the qualitative aspect of color experience is irrevocably subjective (sect. 3.6).

Externalism rejects the move from the possibility of symmetries in the color space to that of functionally indistinguishable color transformations, however, and thus eliminates Palmer's limit on our understanding of color experience. For, according to externalism, the qualitative aspect of experiences of red is not a property of experience itself, but rather is identified with a physical property of physical objects. By ignoring the externalist option, Palmer fails to consider the possibility that we have knowledge of others' color experiences simply by having access to the same colors.

## One machine among many

Barbara Saunders

*Departments of Philosophy and Anthropology, University of Leuven, 3000 Leuven, Belgium.* **pop00127@mail.cc.kuleuven.ac.be**

**Abstract:** In this commentary I point out that Palmer mislocates the source of the inverted spectrum, misrepresents the nature of colour science, and offers no reason for prefering one colour machine over another. I conclude nonetheless that talk about "colour machines" is a step in the right direction.

Though in Notes 1, 4, and 6 Palmer acknowledges the contentious status of the assumptions, he asserts that the colour experiences of humans would be "three-dimensional, would have to include six unique reference experiences (for 'unique colors') at the poles of three axes, would have to include an angular dimension for hue, a radial dimension for saturation, and a linear dimension for lightness, and so forth" (sect. 2.3, para. 8). But if the assumptions are contentious, what basis is there for the imperative? This applies to the inverted spectrum, as well. Treated as unproblematic, it becomes a heuristic to determine "the constraints" on colour.

Locke, Godfather of Empiricism, is presented as the avatar of the inverted spectrum. What Palmer does not seem to know is that Locke generally got his colour ideas from Descartes (see Fodor 1981; Wendler 1996).[1] On Descartes' account, crudely speaking, the spectrum is already in the head and the world is but an arbitrary set of signs triggering the chromatic forms of thought (plenty of room for inverted spectra here). As far as I can see, the unintended consequence of getting Locke into the story is to get Descartes in by the back door – despite the mantle of Empiricism. I suggest Palmer come clean about the real precursor of his colour machine.

Throughout, Palmer's bottom line remains the same. We do not have "even a remotely plausible causal account of how experience arises from neural events" (sect. 3.2), which does not mean that such a theory is impossible in principle, but only that we have not yet hit upon a serious candidate." Even when Palmer is pressed to conclude (given his own Cartesian assumptions) that qualia seem to be beyond the reach of science, there is still the reassurance that "specifying conscious mental states to the level of isomorphism is nothing to be sneezed at." He is aware that assessing "whether our patterns of brain activity are the same is not as straightforward as it might seem." There is a lot we "do not yet know," though "there seems little reason to doubt that someday" we could find out "what particular patterns of neural activity in what particular regions of my brain correspond to my experiences of particular shades of red, orange, green, or any other color." So where does all this modesty and boundless faith in the future get us?



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

Palmer has an idea: Shift authority from troublesome, messy, inchoate, and as yet unknowable "experience" to an eminently knowable simulacrum – a colour machine. What Palmer does not tell us is that the dominant models of colour perception *are already* colour machines – automaton models of colour perception. Thus when he posits "what is known" about the eye, neurons and so forth, *as natural,* he fundamentally misrepresents the state of the art and its relation to *his* colour machine. His colour machine is just one colour machine among many.

Why the preference for this machine rather than another (other than that it captures a particular dominant ideology)? Why not use instead a "Landometer" or "retinex machine"? According to Land (1986), a colour patch in a field of vision can be defined in terms of a triple of numbers representing the three relative lightnesses of this patch as "recorded" by the three types of retinal cones. Using physical luminance as a measure of phenomenal lightness, an instrument can compute the triples of relative lightnesses after screening the whole field of vision, and measuring local lightness on three scales (in accordance with the absorption properties of the three retinal cones). If we were to decide to ascribe to this instrument the capacity "to observe chartreuse" and the other colour shades (putting aside the question of qualia), it is because the results of its computations are first *calibrated* relative to human observations and then proven to correlate well with further human observations. At every stage of the investigation it is manifest *agreement* on what is similar and what is not, that sets the standards (reasoning about animal studies is similar).

Palmer could also say of the idealised Landometer: "The causal isomorphism of its color representations to those of normal trichromats is sufficient to guarantee that it cannot be distinguished from a normal trichromat by behavioral means, but not that it has color experiences of any sort" (sect. 4, para. 5). What distinguishes Palmer's colour machine from the Landometer? Perhaps the general point is that whatever *model* fits the data (whatever the data) it will have more structure and fewer symmetries. With fewer symmetries, fewer "inverted spectra." You can find it all in Descartes, too.

But for the original "sceptical" argument (provided it makes sense at all – which is does not), all this is irrelevant. No matter how many symmetries there are, it is *always* possible to suggest that observed behaviour and *all* scientific measurements underdetermine "raw feel" (i.e., the qualia-aspect) – or to suggest that we are all brains in a vat. If Palmer's colour machine is *at best* one of the *many* models that fit a limited amount of decontextualised data, what is all the fuss about?

The unquestioned acceptance of the "colour-machine" definition of colour and its exploitation by successions of theories presupposes that the processes of "seeing" (of which "colour"-like "space" is taken as a *pars pro toto)* has erroneously been conflated with the scientific method itself. A precursor to the strategy is Newton's setting up of the *experimentum crucis* on the model of the *camera obscura* and of the homunculus role of the spectator (not that I am suggesting this particular variant is still held). The moral is that if colour science were remotely interested in its own history it might learn that it merely repeats itself. From Plato's introextromission theory, to Bacon's rainbow, Descartes' and Newton's prisms, Lockean innate ideas, the Young-Maxwell-Helmholtz versus Hering controversy to the contemporary squabble between the objectivists and subjectivists, nothing much has changed.

Having said this, however, let me conclude on a positive note. I am delighted that cognitive scientists have finally hit on a decent vocabulary. For some time it has seemed downright unfair to go round the world with a machine model of "colour vision" improperly described as "natural" (in the head), fitting putative perceptuolinguistic units to it and slotting them into a set of evolutionary pigeon-holes (e.g., in the Kay & Berlin 1997, pp. 196–201 World Color Survey). It is much more honest to talk about colour machines.

## Computation, levels of abstraction, and the intrinsic character of experience

Jürgen Schröder

*Hanse Institute for Advanced Study, Lehmkuhlenbusch 4, 27753 Delmenhorst, Germany.* **jschroel@urz-mail.urz.uni-heidelberg.de**

**Abstract:** Palmer's color room argument is first contrasted with a different argument by Tim Maudlin against the sufficiency thesis of strong AI. This thesis turns out to be false and hence we need to determine the relevant supervenience base of phenomenal consciousness. That could be done by causal theories and intraindividual experiments. Finally, even if we cannot explain the intrinsic character of conscious states, we may be able to know what the experience of another person is like.

***1. The color room argument.*** To answer whether a material system does have color experiences in virtue of the computations it performs, Palmer adapts Searle's Chinese room to the case of color experience. It is not clear, however, what we should conclude from Palmer's color room thought experiment. On the one hand, he seems to take it as a proof that performing computations of whatever sort and complexity is not sufficient to provide the computing system with any experience. On the other hand, Palmer concedes that there is a straightforward translation of the various objections that have been raised against Searle's Chinese room argument into objections against the color room. For example, the systems reply would hold that the whole system has a color experience, even if the processing part of it does not. Searle tried to rebut this objection by assuming that his processing agent could internalize all the rules and symbols so that he could speak Chinese fluently without understanding a single word of what he himself would say (Searle 1980). At this point the defender of strong AI can say that the agent would either experience understanding or there would be a second stream of consciousness in the agent to which he had no access. Because the status of Searle's argument seems to be problematic, any argument that has the same structure, like Palmer's color room argument, is equally problematic.

Fortunately, there is another argument that has been constructed by Maudlin (1989) and seems much more efficient at refuting the thesis of strong AI concerning phenomenal consciousness. Because his argument is rather complex I will give only its gist and refer the reader to Maudlin's article for the details.

The essential idea is that to perform a nontrivial computation (excluding, for example, computations that assign a constant output to every input), a machine has to have a physical structure that guarantees that the state transitions specified by the machine table would be made if the current input were different, that is, the machine has to have a counterfactual supporting structure. If the structure of the machine does not guarantee these transitions, it does not run the program described by the machine table. Now Maudlin imagines two machines that display the same physical activity during a time interval $t_0$-$t_n$. Only one of them is running a program $\pi$, however, whereas the other is not because it lacks the necessary structure. If we suppose now that phenomenal consciousness supervenes on the physical activity in both machines, then if one has a certain experience the other must have it, too. But according to strong AI only one of them, namely, the one that runs the program, is conscious. So there is a contradiction between the supervenience thesis and the thesis that running a certain program is sufficient for being conscious. The supervenience thesis is not up for grabs for the defenders of strong AI because they need it in some of their own thought experiments (Maudlin 1989, p. 427); therefore, the conclusion must be that the sufficiency thesis is false.

A great virtue of this argument, in my view, is that it separates the question of the adequacy of computationalism from the issue of functionalism. What it shows is that computationalism is an unsatisfactory approach to phenomenal consciousness. It does not show that functionalism is equally inadequate, for it might well be that consciousness is a matter of the pattern of causal interaction



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

in the brain. Two machines may implement the same machine table although the pattern of causal interaction of their physical states may be different. Computation is a matter of a physical activity's taking place in a system that has the right dispositional properties to support a certain machine table.

Having excluded the computational level as too abstract a level for the supervenience of consciousness, one can ask: What is the relevant level of abstraction on which consciousness supervenes? Is it the less abstract level of causal interaction? Or does the causal interaction have to take place in a certain medium, for example, in the medium of electricity? Or does experience even require that certain materials like the ones to be found in nerve cells be involved? It is at this point that the distinction between causal and correlational theories becomes important.

***2. Causal and correlational theories.*** The first task in determining the relevant level of abstraction is to determine which parts of the brain are involved in conscious experience. When these parts are identified we have correlations between activity in these parts and certain types of experiences, for example, activity in a certain part of the visual cortex plus activity in certain parts of the frontal lobes and color experience. As soon as we can be reasonably sure about these locations we could attempt to determine the relevant level of abstraction by systematically intervening into the brain processes occurring at these locations. However, the sort of intervention needed would be much more massive than the application of drugs or the stimulation of nerve cells by electrodes. It would consist of a temporary substitution for brain cells of some other material or a substitution of mechanical energy for electricity. Outside the clinical context where such interventions could restore a certain aspect of experience they would be prohibited for moral reasons.

Now suppose that the relevant level of abstraction was determined and that moreover we know which differences at this level make a phenomenal difference and which do not. We would then have a causal theory telling us who belongs to the same equivalence class of color experiencers. Those people belong to the same class whose brain activities are the same relative to the descriptions picked out by our causal theory.

Nevertheless, this causal theory would not explain why our experiences have the intrinsic characters they have; it would only be able to explain their relational structure. So, in contrast to the case of DNA, the explanatory gap would still be unbridged. The theory would not explain the intrinsic character of our experiences because the connection between a physical activity described at a certain level of abstraction and a structureless intrinsic phenomenal property still appears to be arbitrary. The reason for this appearance of arbitrariness is precisely that the phenomenal property is taken to be atomic or structureless.

***3. Explaining and fixing the quality of an experience.*** There is a certain indeterminacy in Palmer's target article concerning the difference between the explanation and the determination of the intrinsic character of an experience. As just noted, a causal theory does not enable us to explain this intrinsic character. But is it also true that it does not enable us to know what the intrinsic character of an experience of another person is like? The knowledge Palmer is interested in is phenomenal knowledge, knowledge that consists in the capacity to imagine or to have an experience with the same intrinsic character. To have such knowledge one must be able to have the same experience and to have the same experience one must have the same kind of physical activity in one's brain relative to one's causal theory. Because there may be more than one equivalence class of experiencers, the intrinsic character of an experience of an arbitrary person would only be determined for another person if both belonged to the same equivalence class. So instead of the subjectivity barrier that is supposed to exist between any two individuals, we have an "equivalence class barrier," which exists between the various equivalence classes. The latter barrier does not owe its existence to the inadequacy of objective measurement, however, but to the fact that two persons have different physical activities in their brains (again relative to the best causal theory) when they are confronted with the same physical stimulus.

## Consciousness and introspection: How we get to know the inner world

John Smythies

*Brain and Perception Laboratory, Center for Brain and Cognition, University of California at San Diego, La Jolla, CA 92093-0109; and Department of Neuropsychiatry, Institute of Neurology, London, England.*
**smythies@psy.ucsd.edu**

**Abstract:** We can in fact obtain scientific information about the contents of consciousness by the methods of introspectionist psychology. An example comes from the author's work on the stroboscopic patterns and from the way psychedelic drugs alter color perception.

Palmer claims that we can have no scientific knowledge about our experiences, only about the relations between them. Scientific knowledge relating to consciousness and experience, he says, can only be obtained by studying behavior or by studying the brain.

I would disagree. We can obtain valid information about our experiences by examining them (e.g., the contents of the visual field) using introspective techniques under specific circumstances. For example, many years ago I spent two years studying the stroboscopic phenomena (Smythies 1959/1960). These are the geometrical patterns (e.g., grids, checkerboards, families of concentric circles or parabolas, mazes, stars, etc.) that fill the visual field if the retina is stimulated by a stroboscopic light flashing at 4–16 Hz. Using this technique I discovered a large quantity of facts relating to this natural phenomenon that did not depend on either a study of behavior or neurophysiological methods. Certainly I could only study the patterns induced in other people from what they said to me. But I could study my own very similar patterns first hand. At this point I felt it was pointless to give way to philosophical doubts as to whether the patterns seen by my experimental subjects were or were not exactly like or only somewhat like mine. I took more interest in finding out the detail of the forms, colors, and movements of the reported patterns themselves. These phenomena throw light on certain basic aspects of how the visual mechanisms work. The cortical mechanisms convert the intermittent temporal pattern of stimulation received at the retina into complex spatial patterns that fill the visual field in consciousness. Stwertka (1993) has suggested that these patterns originate in the nonlinear (chaotic) networks in the brain. Similar introspective methods now play a large role in contemporary visual science. I do not see the point in claiming that it is impossible, for philosophical reasons, to study the contents of private experience when introspectionist psychologists such as Gregory (1981) and Ramachandran and Blakeslee (1998) are doing just that all the time.

In view of this, I would suggest that there is no need to postulate any basic "subjectivity barrier" or "isomorphism constraint." Phenomenal events simply have the properties that we can observe them to have. I have argued elsewhere (Smythies 1994a; 1999) that, because these phenomenal events that make up the content of consciousness clearly have different properties from the brain events that give rise to them, then, by Leibnitz's Law (Russell 1918), they cannot be identical to these brain events. One rival theory, first put forward by Werner Heisenberg (1958, pp. 106, 179) – the theory of mind-brain complementarity — does not suffer from this difficulty.

Palmer states that we do not currently have a glimmer of a causal theory of how a particular form of brain activity actually produces experience. The theory of mind-brain complementarity would deny that this relationship is causal. Just as an electron may behave either as a wave or a particle, depending on the mode of observation, so a brain may appear as either a collection of neurons or as a collection of sensations, depending on whether it is



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

being observed by a neurosurgeon or by itself. To ask, "how does a pattern of neuronal activity in the brain produce conscious sensations?" is no more legitimate, under this theory, than is the question, "how does an electron as a wave produce the effects seen if we examine an electron as a particle?" I have reviewed other competitive theories elsewhere (Smythies 1994b).

Second, Palmer says that a basic analysis of vision reveals that what we have are experiences of color, for example, "redness." I think it is more accurate to say that we have experiences of colored (e.g., red) patches commonly known as sensations. These may be compounded into visual phenomenal objects as in normal vision, or free of such objects as in the case of after images, eidetic images, and hallucinations. They may also form the "space" and "film" colors that people recovering from cortical lesions experience.

Palmer suggests that pharmacologists might be able to produce drugs that alter colors and cause colors to be experienced as sounds. Such drugs already exist, namely, the psychedelic drugs like mescaline and LSD. They do not turn red into green, it is true, but they do produce supersaturated colors, for example, a red more red than any red normally experienced. Sensory isolation experiments can also induce the experience of "superblack." Psychedelic drugs also induce synesthesias in which primitive sensations are experienced, which are neither visual nor auditory but somehow a compound of both. These drugs induce remarkable changes in shape and movement perception. For example, still objects normally develop complex movements, or a moving object is seen as a series of still ones strung out along the line of motion.

One minor point: Identical twins have far from identical brains (sect. 3.4): For example, they show extensive differences in the pattern of their cortical gyri and sulci.

## Sensory holism and functionalism

Joseph Thomas Tolliver

*Department of Philosophy, University of Arizona, Tucson, AZ 85721.*
**tolliver@u.arizona.edu** **w3.arizona.edu/~phil/faculty.html**

**Abstract:** I defend the possibility of a functional account of the intrinsic qualities of sensory experience against the claim that functional characterization can only describe such qualities to the level of isomorphism of relational structures on those qualities. A form sensory holism might be true concerning the phenomenal, and this holism would account for some antifunctionalist intuition evoked by inverted spectrum and absent qualia arguments. Sensory holism is compatible with the correctness of functionalism about the phenomenal.

I believe that functionalism provides the correct account of the nature of mental states. The functionalist is committed to the claim that mental state properties, such as believing, desiring, intending, or experiencing, are functional properties. Functional properties are relational properties understood in terms of the role their possessors play in the behavior of some system. Functionalism seems an ideal framework for the analysis of the phenomenal properties of color experience, for phenomenal properties appear to be relational properties. The spatial models, which, as Palmer notes, are so apt for the description of color experience, are relational (sect. 1.1 and 1.2). Each kind of color experience is seen as a position in a structure of relative similarity and difference. Having an experience of red is more like having an experience of orange than it is like having an experience of yellow. It is also more like having an experience of pink than it is like having one of neutral gray. For the functionalist, the similarities and differences among types of experiences depicted in the spatial models are similarities and differences among the functional roles of the neural states that embody those experience types.

Palmer offers two objections to this view, one based on the possibility of absent qualia (this is the "color machine" argument and the "color room" argument), and the other based on the possibility of quale transformation or substitution (these arguments are based on the isomorphism constraint). Neither shows that functionalism concerning the phenomenal properties of color experience is false. The mistake is the same in both cases: The arguments assume that the functionalist is committed to the possibility of colors, as we experience them, being defined in terms of the relations of similarity and difference that constitute the relational structure of the color space. The functionalist is not so committed if colors, as we experience them, are not so limited. Perhaps colors are defined not just in terms of their relations of similarity and difference to other colors, but also in terms of their relations of similarity and difference to other sounds, scents, tastes, tactile qualities, and so forth.

In support of this possibility, one can point to the fact that people frequently resort to qualitative metaphors to characterize a quality in some sensory modality. Sounds are said to be high or low (appeal to spatial sensations) or soft and sweet (appeal to tactile and gustatory sensations). Colors can be warm or cool (tactile sensations) or loud (auditory) or vibrant (tactile again). Some combinations of colors can be discordant or inharmonious (auditory). One can dismiss these analogies as mere metaphors conditioned by surrounding cultural or linguistic conventions. But one might regard them as hints at the underlying nature of the qualities themselves. What they hint at is the possibility that both their intramodal and intermodal relations of similarity and difference to other sensory qualities determine the nature of all sensory experience. I call this possibility, "Sensory Holism."

If Sensory Holism is correct, then it is easy to understand why the color machine and the color room do not contain color experiences. They only embody a part of the relational structure relevant to determining the quality of color experience. What makes an experience of red what it is intrinsically includes the way it is similar to and different from an experience of the sound of a trumpet, or the taste of loganberries, or the feel of crushed velvet. According to this version of functionalism, a color machine or color room whose internal states embodied all of these elements of the relational structure of color experience would have color experiences. We need to complicate the examples to determine whether people not in the grip of functionalist dogma would agree that such devices would appreciate colors in just the ways we do.

The intrinsic qualities of experiences are thought by Palmer to lie beyond the "subjectivity barrier." All that a functionalist characterization of those experiences can capture is their relational structure. But individuals whose inner perceptual states have the same relational structure might differ in the intrinsic quality of their perceptual experiences, or their experiences might lack intrinsic qualities altogether. Palmer alleges that we know this because we know that (1) there can be physical differences between persons that make no difference to the relational structure of their experiences; and (2) some changes at the subisomorphism level within single subjects result in detectable changes in the intrinsic character of the subject's experience (sect. 3.5). Of course, what we do not know is whether these two are possible together, that is, whether there are subisomorphic changes that could result in a change in color experience that the person could detect, but let us leave this aside.

I want to focus on the notion of a subjectivity barrier. If there is such a barrier, the subject might also be on the wrong side of it along with the behavioral scientist. It is often accepted, and for good reasons, that in undergoing experiences we acquire more knowledge about the intrinsic quality of our own experiences than anyone else could possibly have, but there are reasons for suspecting that this might be false (or at least less true than we think). The first reason is the commonplace fact that we have such great difficulty in characterizing our experiences and can have only as much success as we do by resorting to the sorts of intermodal comparisons I noted earlier. Consider that the reason this is so might be that each sensory quality is a compound constructed from a limited set of sensory universals that are the basic protophenom-


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

enal[1] units of a combinatorial system that generates all of the sensory qualities. There do seem to be some qualities that are noticeable in many different sensory modalities, including brightness, intensity, and warmth. Many sensory qualities in different sensory domains can be compared with respect to how bright or dark they are. This is certainly the case with colors and sounds. Warmth is a prominent feature of certain tactile sensations, but it is also apparent in certain colors and sounds. Imagine that these cross-modal comparisons are caused by each sensory quality being a structure in which different sensory universals are mixed together in different amounts and in different ways. Imagine also that the sensory universals are never themselves presented in their pure unmixed form. Sensory experience would be made up of a collection of basic elements that are not themselves accessible to introspection. This raises the possibility of variations within a single person over time in the nature of the sensory universals that constitute the intrinsic qualities of his experiences (e.g., differences in the protophenomenal ground of bright/dark variations among his qualities), which would be undetectable by the person himself. The reason I raise this possibility is that I see no reason the sensory universals should not be functionally characterizable, that is, defined in terms of the contribution they make to create a system of sensory differentia.

NOTE

**1.** David Chalmers introduced the notion of protophenomenal properties in his book *The conscious mind* (Chalmers 1996). They are fundamental nonphysical properties that subserve phenomenal properties. My use of the term "protophenomenal" is very different from his. The most important difference is that I do not assume that protophenomena are fundamental in the explanatory order.

## Whatever seems right to me is right

J. van Brakel

*Institute of Philosophy, University of Leuven, 3000 Leuven, Belgium.*
**p6679000@mail.cc.kuleuven.ac.be**

**Abstract:** It is argued that given the task Palmer sets himself, there are no constraints on his colour experiences whatsoever.

Qualia seem to some an intractable obstacle to cognitive science. It is not clear what Palmer's discussions about inverted spectra adds to this worn-out issue (Block 1980; Stillings et al. 1987). At the bottom of Palmer's interest in inverted spectra is his metaphysical belief, nowhere justified, that statements like "we have access to no one's experiences but our own" (sect. 2.1, para. 3) make sense (in the relevant sense of internal experiences being entities with identity conditions). He says (sect. 1.1, last para.): "Because these are private events, any differences between yours and mine can be assessed only indirectly through our publicly observable behavior, as Wittgenstein (1953) argued so forcefully." But what Wittgenstein actually argued so forcefully is that this talk of private events is nonsense, because, in the relevant sense, there are no private events. This way of talking was introduced long ago in philosophy and unfortunately, most scientists (and philosophers under the spell of cognitive science) are still stuck with it. Palmer may say to himself: "In the future I will call only *this* kind of experience." To which Wittgenstein would reply: "Whatever is going to seem right to Palmer is right. And that only means that here we cannot talk about 'right.' Being under the *impression* that you are following a rule (of correctly or incorrectly labelling an experience S) is not sufficient to be *truly* following a rule."

Palmer assumes "that both you and I have the same set of color experiences," and then asks "whether they can be shown to be 'differently arranged,' so to speak" (sect. 1, para. 3). But what is the use of this exercise if there is no scientific or philosophical basis for the assumptions from which he starts? Setting up pseudophilosophical arguments amply spiced with suitably chosen scientific lore may provide enticing brain teasers for readers of *BBS,* but no justification is given that the isomorphism constraint of these block worlds and their epiphenomenal qualia are modelling anything but "self-isomorphy."

Why go with Palmer's metaphysical intuition that it makes sense to say things like "I alone have access to these experiences" (sect. 3.3, para. 4)? Why not follow up on his intuition that "I myself would be hard-pressed to claim, for example, that it seems 'better' or 'more natural' to me that there is a basic color term (BCT) for light reds (PINK) than for light greens, independent of the fact that my language actually has a BCT for light reds and not for light greens" (sect. 1.4, para. 14)? That is an intuition one could build on. Why not also say: "I myself would be hard-pressed to claim, for example, that it seems 'better' or 'more natural' to me that green is a unique hue and not *iban, waln* or *pk,* independent of the fact that I have been brought up with the idea that green is a unique hue." As to *iban, waln,* and *pk* consider Bulmer's (1968) report on Karam "colour" words:

> Leaf surfaces of Munsell rating 5GY/4/4 were variously identified as *mosb* ("dark"), *waln* ("yellow" [roughly co-extensive with *pk*]) and *iban* ("succulent green"), depending on the context of comparison with other leaves, stems or other vegetal parts. . . . As applied to fruit (banana, papaws) it [i.e., *pk*] covers the Munsell ratings approximately 5YR-5Y6-8/7-8; as applied to human skin, approximately 2.5-5YR4-5/4-6, which would normally be described as *gs* ["dull brown, green, or olive"] or *gac* ["dirt or mud"] in the case of fruit skins.

If we like the sort of games played by the friends of qualia and inverted spectra we can also play it in the language of the Karam people. How do I know (being a Karam speaker) that another Karam speaker's experiences of *iban* and *pk* are not reversed compared with mine? We might both space *waln* and *pk* on equal distance on some psychologist's just noticeable difference (jnd)-scale, but what does this say about our true experiences of *waln* and *pk*? How would Palmer's account of distinguishing the nature of colour experiences from their structural interrelations apply to Karam "colours"? Already raising the question strikes me as nonsense and probably Palmer would agree, because he might consider quoting such esoteric examples utterly irrelevant.

So let us stay home and take for granted that colour is colour and that is what we should be talking about. Let us ask what local support there is for one of Palmer's more mundane assumptions. Consider his description of the colour circle with red/green and yellow/blue at opposite poles of orthogonal diameters. This seems plausible to Palmer (and many of his readers). But why? The colour circle in Newton's *Opticks* (1952) does not confirm these insights. But there is probably no reason to trust people like Newton. There must exist some sophisticated experiments that explain why we should prefer Palmer's colour circle over that of Newton. And there are. Hardin, referring to a proximity analysis of qualitative similarity judgments of colours says (1988, p. 42): "Notice particularly that the unique hues . . . are spaced about 90 degrees apart." However, the picture he presents would fit the painter's colour wheel better (with the primaries yellow, blue, and red placed 120° apart). In the colour hexagon used by printers, 6 primaries (magenta-red, violet-blue, cyan-blue, green, yellow, and orange-red) are "naturally" placed 60° apart – a variant of the numerous colour circles from the past that had as opposites red/green, blue/orange, and yellow/violet (Gage 1993). Indow (1988) provides data spacing the 5 Munsell primaries red, purple, blue, green, and yellow about 72° apart. Why is it not equally plausible that the Munsell system, which allegedly is based on jnd's, confirms that the five primary Munsell colours are equally spaced? All these representations of the "isomorphy constraints inherent" in the colour circle look equally "natural." Perhaps it is wiser for Palmer to conclude, given the task he has set himself, that there are no constraints on his colour experiences whatsoever.



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

# Out of sight but not out of mind: Isomorphism and absent qualia

Robert Van Gulick

*Department of Philosophy, Syracuse University, Syracuse, NY 12344-1170.*
**rnvangul@syr.edu**

**Abstract:** The isomorphism constraint places plausible limits on the use of third-person evidence to explain color experience but poses no difficulty for functionalists; they themselves argue for just such limits. Palmer's absent qualia claim is supported by neither the Color Machine nor Color Room examples. The nature of color experience depends on relations external to the color space, as well as internal to it.

Palmer proposes the isomorphism constraint as a limit on what behavioral evidence can show about the nature of color. Third-person data, whether behavioral or neural, can tell us about the relational structure of color experiences but not about their intrinsic natures. Insofar as alternative isomorphic sets of experiences can have the same relational structure, behavioral evidence can impose no finer constraint.

Palmer views this result as raising an objection to functionalism, which aims to classify mental states in terms of their functional roles. Given the isomorphism constraint, the intrinsic nature of experiences seems to escape the functionalist's net. He writes:

> These nonrelational aspects of experience lie, by definition, outside the domain of functionalism; they are underconstrained by relations among mental states. It seems that the failure of functionalism to provide an account of these aspects should be counted against its claim of fully specifying the nature of mind. It does not seem to be able to do the whole job. (sect. 4)

As a functionalist – whether "card-carrying" or not (sect. 2.4) – I find this "objection" puzzling. I agree that behavioral evidence can reveal only the relational structure of color experiences and not their intrinsic character. But I do not see that as an objection, indeed it is a bit of common wisdom among functionalists, going back a quarter century or more (Shoemaker 1975). Functionalists also long ago accepted the possibility of "alien qualia," that is, color experiences totally distinct from our own but sharing the same relational structure (Shoemaker 1981). I suppose some hypothetical functionalists might be taken aback by their inability to do the "whole job," but most actual functionalists have been well aware of just such limits for a long time. Moreover, some of them go a step further to show that a functionalist theory of understanding predicts and explains just such third-person limits on the effability of intrinsic qualitative character (Van Gulick 1985; 1991). If so, rather than posing a problem for functionalism, such limits provide further confirmation of its truth.

There is, however, an issue about which Palmer and functionalists surely disagree: the possibility of absent qualia or color zombies. Palmer moves beyond the possibility of inverted qualia to the stronger claim that two systems, only one of which has experiences, might nonetheless share all their relevant relational structure (sect. 2.4). The weaker (inverted) possibility does not by itself entail the stronger (absent) possibility (Shoemaker 1975; 1981); additional argument is needed. Palmer imagines a "color machine" that processes light as our human visual system does (at least to the level of computational sameness) and responds to colored stimuli as people typically do. He argues that the machine's relational structure would be isomorphic to that of our own color experiences. Thus he concludes that relational facts alone cannot exclude the possibility of color zombies. He also asserts that functionalists (at least "card-carrying" ones) "would claim that such a machine does have color experiences purely by virtue of the computations it performs," a claim he characterizes as seeming "unlikely to readers not in the grip of functionalism" (sect. 2.4).

However, few if any functionalists would regard the color machine as having experiences, nor does anything about the theory support such a view. The machine may share some computational structure with our visual processing system and perhaps some links to a "verbal" output system, but no one supposes that those modules suffice to produce a system with conscious experiences of color or of anything else. Any plausible functional model of an experience will be far broader and more holistic. Visual experience occurs as the experience of a conscious subject or self perceiving its place in a world of visually presented objects. It is likely to require the globally integrated activity of many different systems throughout the brain. How the outputs of various sensory pathways get integrated into a conscious percept is as yet not well understood, but it is a subject of intense neurophysiological investigation and offers hope for the sort of causal theory of consciousness that Palmer contrasts with merely correlational ones.

Whatever the outcome of those empirical studies, the philosophical point is clear. Functionalism from its inception has been holistic in its view of mind and especially of consciousness. Thus the imagined "color machine" does not even come close to satisfying the conditions a functionalist would find plausible. Nor would the machine's verbal output satisfy the functional conditions to count as "naming," "judging" or "agreeing." Despite its superficial verbal similarity to our responses to the same stimuli, the machine's outputs lack the broad context of other behaviors (both verbal and nonverbal) within which they would need to stand to count as genuine speech acts of the relevant sort. Contrary to Palmer's claim, no functionalist would regard the color machine as a subject of color experiences.

The same point applies *mutatitis mutandis* to his Color Room example (sect. 2.5). Palmer is right that satisfying the conditions in his example would not suffice to produce color experiences. But he is wrong to claim that performing the limited range of operations in the Room would "mean that you have satisfied the usual functionalist criteria for claiming that you have . . . color perception and naming" (sect. 2.5). Here, too, considerations of holism and context would make those limited actions far from sufficient for a functionalist.[1]

The take-away moral would seem to be that one must consider both external and internal relational structure. Given his concerns about the possibility of inversion, Palmer focuses on intraspatial relations within the color space. But the relations that the color space bears to other organized spaces within our mental and representational domain are just as important to understanding its role in conscious experience. We do not simply experience colors. We experience a world of objects with colored surfaces presented to us at a given time, place, and context within a meaningful world. Sometimes when philosophers talk about the phenomenal nature of experience they use that word interchangeably with "qualitative" and refer only to the supposed raw feels of experience. That is one legitimate way to use the word, but there is an alternative use that owes more to Kant and to phenomenology that emphasizes the globally meaningful organization of the world as we meet it in experience of the phenomenal world. Any functionalist or relational theory of consciousness will need to take account of the larger overall context if it is to understand color experiences as color *experiences.*

NOTES

**1.** Confusion may arise from the fact that when we explain, we do not feel the need to spell out all the assumed background context. If I say you can fly to California tomorrow if you have an open return ticket, I do not need to add that of course you need a functioning airline with planes, pilots, check-in agents, and a running computer system willing to accept your open ticket. You do need all those things, but we do not feel the need to list them. Similarly, a functionalist discussing one particular aspect of mind (e.g., color naming) may focus on the specific features most relevant to that particular feature without bothering to mention that those conditions must be fulfilled within a much larger context of organization and behavior (e.g., those necessary for being a genuine language user able to name things at all.)



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

## The possibility of subisomorphic experiential differences

Christopher D. Viger

*Center for Cognitive Studies, Tufts University, Medford, MA 02155-7059.*
**cviger01@emerald.tufts.edu** **ase.tufts.edu/cogstud/mainpage.htm**

**Abstract:** Palmer's main intuition pump, the "color machine," greatly underestimates the complexity of a system isomorphic in color experience to humans. The neuroscientific picture of this complexity makes it clear that the brain actively produces our experiences by processes that science can investigate, thereby supporting functionalism and leaving no (color) room for a passive observer to witness subisomorphic experiential differences.

Palmer distinguishes simulating color experiences from actually having them by imagining a "'color machine' that actually processes information from light in the *same* way as people do and that responds *as people typically do.*" (sect. 2.4, para. 3, my emphasis). Palmer's thought experiment relies on his intuition that "it would not be very difficult" (sect. 2.4, para. 3) to construct such a machine. Palmer even offers a schematic; a few mirrors, prisms, cardboard masks, photocells, and computational circuits are up to the task. But the plausibility of such a construction being sufficient derives from grossly underestimating the task. People respond to colors with incredible subtlety and variety, including changes of mood, preference selections, and aesthetic judgments. On Palmer's own terms the "color machine" must respond likewise. But once the task is made clear it is also clear just how difficult it would be to construct such a machine, as any AI researcher will attest. All Palmer offers is the "front end" of such a machine, which might indeed not be too difficult to construct – but no functionalist would declare that our eyes and, say, V1 (the "front end" of our visual system) have experiences. It is only by various mechanisms playing a functional role within a very complex system that experience can arise at all, and in such cases it is the *entire system* that is the experiencer. Only by presupposing qualia is there any motivation to attribute experiences to subsystems, which is the flaw with the color room thought experiment. The subsystem in the color room is not the locus of experience. Once we expose the requisite complexity for a system to be experientially isomorphic to humans vis-à-vis color experience, the force of Palmer's thought experiment evaporates. *Of course,* the kind of machine he imagines would not have color experiences. It has nowhere near the complexity required to have experiences of any sort.

Consider the neuroscientific account of coming to have a color experience. "When an individual cone absorbs a photon, its electrical response is always the same, whatever the wavelength of the photon" (Kandel et al. 1995, p. 456). Individual cones are more likely to absorb photons of particular wavelengths depending on the pigment they contain, but the response of an individual cone does not tell the brain anything about color, because the response is stochastic. It is only the comparative strength of the responses by each entire subsystem of cones containing a particular pigment, subsystems large enough to exploit the stochastic responses of individual cones, by which color can be discriminated. "For example, if an object reflected primarily light of a long wavelength, the response in the longer-wavelength cone system would be stronger than the response in the other system, and higher processing centers would interpret the object as being red or yellow" (Kandel et al. 1995, p. 458). What emerges from this neuroscientific picture of color perception is that the brain *interprets* objects as having a certain color; our brains actively produce our color experiences. "Vision is not merely a matter of passing perception, it is an intelligent process of active construction" (Hoffman 1998, preface, p. xii). It follows that experiences admit no subisomorphic experiential difference, because the brain's constructions are not the privileged possessions of the person whose brain is doing the constructing. For example, consider Palmer's case of reversing pigments in the cones.

Palmer supposes that someone whose L-cones have the M-pigment and whose M-cones have the L-pigment would be "red-green reversed trichromats" (sect. 1.3, para. 5). He notes, however (n. 5), that red-green reversed trichromats could be the case only if there were some difference other than the pigments that L- and M-cones contain, otherwise all that would be reversed are the L- and M-cones themselves. But such a difference would have to be a functional difference, in that the distinct subsystems containing a specific pigment would have to play different functional roles within the larger nervous system. Then, because the brain interprets color based on the relative strengths of responses from each subsystem of cones containing a particular pigment, a strong response by a particular subsystem would be interpreted as a certain color, *in virtue of the functional role of that subsystem.* In such a case reversing the pigments in the cones, but not the functional roles of the subsystems of cones, a red object would produce a strong response in the subsystem of cones that results in the brain interpreting the object as green. So the red object would be experienced as green. But it would also be reported as green, because the functional role of the subsystem leads to the object being interpreted as green, breaking the isomorphism between the original trichromat and the red-green reversed trichromat. If we further suppose that the red-green reversed trichromat would report the object as red and act behaviorally identical with someone who experienced it as red to preserve the isomorphism, we would also have to reverse the functional roles of the subsystem of cones. (See Dennett 1991, pp. 389–98 for a more detailed discussion to the same effect.) The point is simply that experiential differences that Palmer supposes to be subisomorphic actually presuppose functionalism to be coherent, in which case they are behaviorally detectable. The relational structure of experiences, up to isomorphism, entirely captures our experiences.

Paradoxically, what Palmer and other defenders of qualia require are internal states that could make no difference – *no difference* – not to emotional states, aesthetic judgments, preference selections, or any of the myriad other ways that our color experiences matter to us. But if they can make no difference, it should make no difference to the defenders of qualia if there are no such states.

## Isomorphism: Philosophical implications

Edmond Wright

*Hon. Member SCR, Pembroke College, Oxford; 3 Boathouse Court, Trafalgar Road, Cambridge CB4 1DU, England.* **eew20@hermes.cam.ac.uk**

**Abstract:** The originator of the notion of structural isomorphism was the philosopher Roy Wood Sellars. Many modern philosophers are unaware how this notion vitiates their attacks on the concept of an internal sensory presentation. His view that this allowed for corrective feedback undercuts Palmer's belief that there is a mapping of objects. The privacy of subjective experience is also shown not to be inviolable.

Palmer's target article is welcome, especially in view of the fact that a number of philosophers resistant to acknowledging the existence of experience internal to the brain have not taken account of structural isomorphism. To take one example of many, John McDowell, in his recent *Mind and world,* repeatedly rejects the notion of resemblance between a non-colored light-ray input and an internal sensory matrix. His rejection, however, takes the form of merely saying that the claim is incomprehensible: "How can there be a resemblance between a color and something we can't characterize in terms of how it would look?" (McDowell 1994a, p. 113). He is thus not aware of a *nonpictorial* resemblance based on an isomorphic match, preserving only relations across a field, precisely what Palmer is putting forward here. Roy Wood Sellars, probably the first to argue for isomorphism between input and sensory presentation, which he called "structural similarity," wrote in 1922 that the only copying is of relations, mediated by causal ratios (Sellars 1922,



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

p. 37). His view was discussed by Chisholm (1954), though without seeing the further implications, and taken up later by Wright (1986, pp. 13–14) and Maund (1993, pp. 57–58) to sustain a case for internal sensory presentations. Sellars, like Palmer, was well aware that he was thus distinguishing himself from the traditional empiricist Sense-Datum Theory, for he points out that Locke "did not see the possibility of similarity between cause and effect on the lines of a reproduction of pattern" (1932, p. 111).

Here, from Sellars' comment, arises the first qualification of Palmer's argument. Palmer, having correctly accepted, with Sellars, that it is a pattern, *a structure of intensities,* that is transferred, moves without explanation to its being *a mapping of objects,* for he says: "Preserving relational structure appears to be a necessary condition for one set of objects to represent another" (sect. 2.2, para. 5). He also asserts correctly that biological considerations support the view that color experiences differ from person to person (sect. 2.3, para. 6), and that just noticeable differences (jnd) and range sensitivities are measurably rarely the same (whatever the sensory mode); he even allows with Gregory (1983, p. 451) that there could be a sound/light interchange. What has not occurred to him is that, given this latter assertion of differences of input, it is not possible for each person's selection from that input, that is, what is to constitute "an object" for that person, not to differ also. There is the problem of coordination across persons to consider, one that I have argued is conducted by each agent by joining in a mutual assumption with his or her opposite in the moment of communication that *there already exists one entity on which they are converging in understanding* (Wright 1990, 73–75). As both the sociologist Alfred Schutz (1962) and the psycholinguist Ragnar Rommetveit (1974; 1983) have independently maintained, we create a *partial* objectivity by behaving together as if we had achieved a *perfect* objectivity. This is of the utmost evolutionary value, because we can capitalize on our differences of sensing and perceiving to update and (we hope) improve the so-called common concept. There is actually no practical necessity that there exist in the real, so to speak, "a single object" at all, merely a region of intensities on which a rough converging coordination has been achieved. To use Harnad's terms, there is a never-to-be-achieved perfect "convergence" of each agent's "approximations" to the putatively common categorization (Harnad 1987, p. 538).

There is thus a social parameter to consider that takes advantage of the fact that we are each provided with a differing field of intensities on which, either through direct motivation or by the motivation that is transferred to us in communication, we can move our gestalts about to alter what "color patches" – though I would prefer to say three-dimensional regions – we are, for the time being, *to take as* objects. The "mapping of objects" is thus part of a intersubjective venture in getting the most viable sorting out of those regions of our sensory presentations, the intensities of which are our only guide to the mass/energy world outside (and bodily – inside) us. This is, of course, a version of a naturalized, genetic epistemology, because it claims, with Piaget, that, because assimilation and accommodation are continuous, objectification is always no more than viable (Piaget 1970, p. 15; also see Hooker 1995, pp. 257–67; von Glaserfeld 1982, p. 613).

Given this approach, which seems to be the most likely conclusion to be drawn from Palmer's argument, then Palmer ought not to regard the Zombie Objection as one with any force. Zombies, like digital robots, work on "common" concepts, the criteria of which are made up from a list of individually unalterable elements. They, like the person in Searle's Chinese Room, would be quite unable *to change the language for all those who are using it.* It is precisely because we are operating with a field of intensities, structurally similar to the external input, that we are able to re-jig our percepts and concepts. It is not objects that are mapped at all: We do the mapping, that is, the choosing of the most viable selections and treating them as "objects." If they were directly mapped, we should not be able to change our selections when something rewarding or aversive occurs. This bears out Sellars' emphasis on feedback (1969, p. 142).

Because this view takes the sensory fields as nonepistemic, as knowledgeless evidence on which the motivation system works to select percepts, there is no reason why neurophysiology should not one day discover how they are created, nor why it should not be possible to connect mind to mind. Palmer seems unduly restricted by a lingering Cartesianism in his claim that sensory fields must remain wholly separate. Why should it not be possible to make a cross-connection and share even a minuscule part of another organism's field? There is no "constraint" at all.

# Author's Response

## On qualia, relations, and structure in color experience

Stephen E. Palmer

*Department of Psychology, University of California, Berkeley, CA 94720-165.*
**palmer@cogsci.berkeley.edu** **socrates.berkeley.edu/~plab**

**Abstract:** In this Response, I defend the notion of intrinsic qualities of experience, discuss the distinction between relational experience and relational structure, clarify the difference between narrow and broad interpretations of color experience, argue against externalist approaches to color experience, defend the concept of isomorphism as a limitation in understanding color experiences, examine critiques of the color machine and color room arguments, and counter objections to within-subject experiments based on memory limitations.

To begin, I would like to thank all the contributors for the time and effort they have invested in their commentaries on my target article, "Color, consciousness, and the isomorphism constraint." I have learned a great deal by reading and considering the varied remarks, even though I have usually come to the unsurprising conclusion that I disagree with them when they disagree with me. In response to criticisms, I have tried at least to clarify the issues and to restate my own views more carefully and coherently than in the target article.

Because I do not have the space to respond in detail to every commentary, I have chosen to respond according to a number of general issues described by the main headings. There are also two commentaries to which I will not respond: **Block**'s because the issue he discusses (representationism), while interesting, is not one I addressed in the target article, and **Brill**'s because I do not see any fundamental disagreement with me in his photometer discussions. I also will not comment on **Smythies**'s reference to Heisenberg's theory of mind–brain complementarity because I do not understand this theory. About all the rest, however, I will have at least something to say.

The central claim of the target article is that a "subjectivity barrier" divides what can and cannot be shared and communicated between people about the nature of their experiences. I further claimed that this "metaphorical brick wall" separates what can be known by objective scientific means from what cannot. I then identified this limitation with the relational structure of isomorphism: Anything below the level of isomorphism lies purely within the subjective domain and cannot be known with certainty by others or by objective science. No functionalist account


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

can reach down to the subisomorphic level to explain the intrinsic nature or even the existence of different experiential qualities.

**R1. Intrinsic qualities of experience.** My claims about intrinsic qualities of experience (or "qualia," as philosophers tend to call them) provoked many commentators to challenge a variety of issues from a variety of viewpoints. Some question the logical coherence of my conception of intrinsic qualities of experience in the first place (e.g., **Myin**), others simply deny their existence (e.g., **Dennett**), and still others deny only that there could be any experiential *differences* between observers at that level (e.g., **Viger**). I will respond to each objection in turn.

**Myin** argues that my description of intrinsic aspects of experience simply does not make sense. If the intrinsic qualities of experiences are merely what is left over after subtracting out all the relational aspects, then intrinsic aspects of color experiences cannot have any relations to each other. His objection is that there is little, if anything, left to "orangeness" once all its relations to other color experiences are removed.

I believe the confusion underlying this argument is a subtle one concerning the difference between *relations* and *relational structure.* I was not as clear about it in the target article as I should have been, at least partly because I am not as clear about it in my own mind as I would like to be. Since **Myin** has called me on it, however, I will attempt to clarify my thinking.

I tried to be careful in identifying what is knowable to others and to science through behavior as the *relational structure* among experiences, not the relations themselves. Relations among experiences have intrinsic qualities too – ones that are intimately tied to the intrinsic qualities of the individual experiences they relate – and these relational aspects of experience are just as unknowable as individual aspects of experience. There is something it "feels like," shall we say, to experience the stark contrast between adjacent white and black regions, for instance, even though this contrast is a relational aspect of experience. Moreover, I cannot know whether what I feel on viewing this white/black contrast is the same as what you feel because the metaphorical brick wall is very much in place for the intrinsic quality of this relational experience. So there is lots of relational stuff on the "intrinsic qualities" side of the subjectivity barrier (i.e., below the level of isomorphism); it has not been "subtracted out" when we distinguish relational structure from intrinsic qualities. This means that the "orangeness" of experience on viewing an orange is still all there when I speak of its intrinsic qualities.

So what, then, is the nature of this other "structural" stuff? I believe this question to be a deep and poorly understood one in cognitive science – mathematicians seem to have a clearer idea – but I will take a shot at it anyway. My own current views are perhaps most closely related to those of logician Alfred Tarski (1954), cognitive psychologist Wendell Garner (1974), and measurement theorists Krantz, Luce, Suppes, and Tversky (1971). As I use the term, structure consists of constraints on a set of elements (or objects) imposed indirectly on those objects via relations between two or more of them. The individual objects themselves are completely unconstrained, but relations among them are not, and these relational constraints constitute their structure (or equivalently, their relational structure, since there is no structure at the level of individual elements).

To illustrate, consider the case of perceived lightness of achromatic surfaces. The "objects" are tokens of color experiences that arise from viewing gray surfaces from white to black. The structure among these objects is the set of constraints induced by the "lighter than" relation for ordered pairs of gray-scale experiences. This structure consists in the fact that the lighter-than relation only holds for certain, nonrandom, ordered pairs of gray experiences, that is, there is systematicity in which ordered pairs are and are not related in this way. Specifically, if A is lighter than B, then B cannot also be lighter than A. This constraint, called *antisymmetry,* is part of the structure of the lighter-than relation. Another constraint, called *transitivity,* also holds for lighter-than: If A is lighter than B, and B is lighter than C, then A must also be lighter than C. These constraints do not hold for all relations (e.g., "has the same lightness as" is symmetric rather than antisymmetric), but together with others, they determine the structure of the lightness dimension. (Notice that they do not – indeed, I would claim, cannot – define the lightness dimension itself, but only its structure, because the same linear dimensional structure characterizes many other sensory dimensions, such as "bigger than," "louder than," and "saltier than"). It is this conception of structure that underlies an isomorphism between two sets of experiential objects.

The picture that emerges is not one of individual experiences on the subjective side of the brick wall and relational experiences on the objective side (as **Myin** interpreted my distinction), but of experiences (including relational ones) on the subjective side and experiential structure on the objective side. Notice that the special importance of relational experiences is only that they determine structure, not that they are magically on the objective side of the brick wall. Moreover, intrinsic qualities are not what is left over after the structure has been "taken out," because the structure isn't actually taken out so much as "mirrored" or "abstracted" in the objective domain: Structure exists on both sides of the subjectivity barrier. The situation is very much like that in formal mathematics, in which there is a domain of objects that has some structure, and mathematicians attempt to capture that structure at an abstract level that does not depend on the identity of the objects. The objects and their relations have some determinate structure, and mathematicians formalize it in a very general way. (See the replies to **Harrison** and **O'Brien & Opie** for discussion of related topics.)

**Dennett** does not deny the coherence of intrinsic qualities of experience, but he does claim that people have no access to them.[1] He doesn't actually say that such experiences don't exist, but he does claim that we only have access to the *relations* we detect between these inaccessible experiences. This is, of course, perfectly consistent with functionalism because functionalism is based on causal relations among mental states. It is an interesting position in part because relational information is so important in determining experiences in many domains, color included, that one can almost get away with it. Dennett didn't bother to cite any supporting evidence for his claim, but in the interest of proper appreciation for this possibility, I will do it for him.

The perception of color and lightness depends crucially on spatial and temporal relations or contrasts (i.e., edges



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

and/or gradients in colored space-time). Even a surface that initially appears to be strongly colored will eventually lose this appearance if it is uniform in color and texture over the whole visual field (i.e., a *Ganzfeld*) and is viewed for an extended period of time. The visual system eventually adapts to the point that the surface appears neutral gray, no matter what color it appeared initially. One reasonable account of such effects is that they result from the lack of relations to any other color in space or in time. This kind of "relational determination," as the Gestaltists called it, is not restricted to Ganzfelds either. Wallach's (1948) classic studies of lightness perception and Land's later retinex theory (Land & McCann 1971) both highlight the fact that the visual system is primarily sensitive to contrast ratios between the luminances of adjacent regions. More complex effects of lightness constancy that result from the interpretation of retinal luminance edges (environmental intensity edges versus reflectance edges; see, e.g., Gilchrist 1988) are also likely to be compatible with relational approaches, because all the information required to interpret the edges is embedded in the relations among the image luminances.

The problem is that although relational effects are crucial in determining visual experience, they are not the whole story. We perceive not only relations, such as contrasts and luminance ratios, but also determinate colors of individual surfaces. In lightness perception, this is called the "scaling problem" because once all the appropriate ratios are known, it is still unclear how the entire structure of ratios maps onto the absolute lightness scale from white to black. One rectangle of a Mondrian display looks white, another looks medium gray, and a third looks black; they do not float unanchored in a sea of ratios. Even in the case of a Ganzfeld, the visual field always has a determinate color appearance. First it is, say, yellow, and it gradually fades to gray, but it does not *disappear,* as it seems it should if we experience only relations to other regions adjacent to it in space and time. The rules by which contrasts and ratios are anchored to perceived lightness are beginning to be understood, at least in the restricted world of achromatic color perception (e.g., Cataliotti & Gilchrist 1995). One rule is that, with proper experimental conditions, the lightest region in the entire visual field is generally experienced as white, and other regions are scaled in lightness relative to it, as determined by their ratios to the lightest region.

As interesting as **Dennett**'s claim is, I believe that it is fundamentally mistaken and possibly incoherent. Even by his own hypothesis, there must *be* intrinsic qualities of experience, for otherwise there could be no relations between them to be detected. But relations are attributes defined only for two or more objects. If we have access only to relations between experiences, how is it that we experience qualities of individual objects? The lightest rectangle in an achromatic Mondrian looks *white,* not just twice as light as the next-darker rectangle. Dennett's view simply does not square with either introspective experience or with scientific findings.

**Viger** does not deny that individual experiences have any intrinsic qualities, but he doubts that there could be any difference in experience that would occur at this subisomorphic level. Specifically, he argues that pseudonormal vision (i.e., switching the pigments in the M-cones and L-cones) would not produce any change in experience because there would be no change in the functional roles of the subsystems, and it is only their function that matters. He therefore concludes that pseudonormal vision poses no threat to functionalism and, moreover, asserts that I actually presupposed functionalism in my argument against functionalism via pseudonormal vision. Both of these claims are mistaken.

First, it is **Viger** who presupposes functionalism in his argument, not I. He simply asserts that the brain interprets seeing a certain color based on a strong response by a particular subsystem, *in virtue of the functional role of that subsystem* (his emphasis). I make no such claim, as I will show by describing a simple example of how color experiences could plausibly be reversed by switching the photopigments in M-cones and L-cones.

I begin by assuming that the color a person experiences is determined by the firing of particular genetically determined central neurons, irrespective of their "functional roles" in the system. Since we only need to concern ourselves with red and green experiences at one spatial position for pseudonormal vision, we will simplify matters and refer to the single central neuron whose firing causes red experience in that position as the "R-cell" and the one whose firing causes green experience as the "G-cell." The central R-cell is prewired to receive activation from the retinal receptors that usually contain the L-pigment (i.e., L-cones) and to receive inhibition from certain other retinal receptors that usually contain the M-pigment (i.e., M-cones). The central G-cell has the opposite preprogrammed connections from M-cones and L-cones. Predominately long-wavelength light will thus cause the R-cell to fire vigorously, but not the G-cell, and medium-wavelength light will cause the G-cell to fire vigorously, but not the R-cell, at least for color-normal observers.

Now suppose that in pseudonormal individuals, genetic mutations that are quite independent from the genes that control the wiring from M-cones and L-cones to the R-cell and G-cell cause the L-cones to contain the M-pigment and the M-cones to contain the L-pigment. The result will be that the R-cell will receive excitatory input from L-cones containing M-pigment and inhibitory input from M-cones containing L-pigment; the G-cell will receive excitatory input from M-cones containing L-pigment and inhibitory input from L-cones containing M-pigment. The net effect in pseudonormal observers will be that predominately long-wavelength light will cause the G-cell to fire vigorously, and medium-wavelength light will cause the R-cell to fire vigorously. By the original assumption that experienced color is determined by which central is firing, light that appears red to a normal observer will appear green to the pseudonormal observer, and vice versa.

Regardless of whether there actually are any pseudonormal observers and whether the assumptions I have made are empirically correct, I do not see any logical problem with this scenario.[2] It is not a functionalist scenario because of the initial assumption that color experience is determined by which particular central neuron is firing, regardless of the pigment in the receptors to which they are connected. The crucial factor is that the difference I have assumed between red and green internal experiences is not determined by functional roles within the system, but by genetic designation. (This may depend on the functional roles of such cells for the species as a whole, but not on their roles within the individual.) I make no claim that this is empirically correct, but I do believe it is coherent and that it straightforwardly predicts subisomorphic experiential differences between normal and pseudonormal observers.


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

**R2. Narrow versus broad views of color experience.** Another basic issue that underlies many of the commentaries is the question of what counts as color experience. In the target article, I took the position that it was a relatively circumscribed aspect of phenomenology that surely included primary sensory appearances derived from different physical spectra of light and the sensory relations between such appearances, but not much else. The reader will recall that I was even circumspect about including color-naming categories, such as basic color terms (BCTs), as being truly part of color experience. I did end up considering them on the grounds that systematicities in BCTs might indirectly reflect subtle inhomogeneities in sensory experiences that are otherwise difficult to substantiate. Let's call this "the narrow view" of color experience.

Many commentators (e.g., **Frumkina, Myin, Tolliver, Van Gulick,** and **Viger**) disagree strongly with this narrow view in one or more respects. They argue that color experience also includes one or more of the following: cross-modal sensory components, chromatic mood induction, memories associated with colors, aesthetic preferences among colors, and salient cultural conditions. One problem with this strategy is: Where (if anywhere) does one draw the line in deciding what is part of a color experience? If one includes aesthetic responses to the color being viewed, what about the aesthetic responses to its combination with other colors? Or with all possible triples or even n-tuples of other colors? If one includes associative memories (e.g., the ripe tomato that I most closely associate with the precise shade of red in question), are we to include the shape of that tomato in the experience of that color? What about nonvisual aspects, such as its taste and smell? Or perhaps the sequence of phonemes that constitutes the name of the company that made the television on which I saw the image of the bottle of catsup that had the picture of the tomato that exemplified the precise shade of red in question? This "broad view" of color experience does not seem like a productive approach, at least to me, as it will count every aspect of experience as part of color experience.

Many of the commentators who take this tack seem to be motivated by trying to "save" functionalism in one way or another. One rationale is that even if there are a few underlying symmetries in the narrow view of color experience I advocate in the target article, expanding color experience to include even a few of these other aspects of experience surely breaks every possible symmetry, and this seems to save functionalism from the potential predicament posed by Locke's argument. Even so, expanding color experience does not affect either the alien color argument or the color zombie argument because neither assumes interobserver automorphism of experiences. That is, even if all symmetries were broken by broadening the nature of color experience, this would not block replicating the same relational structure via an isomorphism to a different set of experiential dimensions or to a set of dimensions devoid of experience (see sect. R4).

Other commentators sought to broaden the domain of color experience for somewhat different purposes. They argue that the additional complexity of such expanded color experiences refutes the color room argument. I will consider this argument in section R5 below.

I am not claiming that the narrow view of color experience I have taken is "correct" and that the broad view is "wrong." Trying to partition experience into different aspects is a difficult enterprise at best, and I do not want to imply that my narrow view is unproblematic. I took it in part to provide the greatest likelihood of supporting the unexamined, commonsense "yes" answer to the question of whether you and I have the same color experiences under the same external conditions. If moods, aesthetic preferences, and associative memories are included in our conception of color experience, then the answer to this question is surely "no," but for what seem to me less interesting reasons: We all have different memories, associations, moods, preferences, and the like due to our different life histories. To the extent that these are included in the definition of color experience, they preclude the possibility of color experiences being the same across people. The broad view of color experience thus throws out the baby with the bathwater.

Moreover, even after considering the commentator's objections, I believe that there is a sensible notion of color experience – perhaps I should call it "basic" or "pure" color experience to differentiate it from the more complex and inclusive notion some of the commentators argue for – that does not include preferences, associative memories, moods, linguistic labels, or even basic color terms. This is the sense in which the same bowl of, say, split-pea soup looks the same color to someone who adores that particular shade of green, someone else who despises it, someone who eats pea soup daily, and someone who has never seen the stuff. This is the narrow notion of color experience I intended to discuss in the target article.

Even so, I do have a serious concern, mentioned by **Frumkina,** about how representative this situation might be for other forms of sensory experience. I personally find it much harder to separate aesthetic responses from an underlying "pure" sensory experience for tastes and smells. Being strongly liver-phobic, for example, I cannot experience its taste or smell without having an intense liver-specific revulsion. Is there a more basic taste or smell of liver for me that strips off the disgust? Maybe so, but the answer is far less clear to me than in the case of color.

**R3. Internalist versus externalist approaches to experience.** The commentators **Byrne, Ross,** and **Malcolm** object that my conclusions in the target article are faulty because I take an "internalist" approach to color experience. They advocate instead an "externalist" or "objective" approach in which my color experiences, like the perceived redness of a rose, are identified with a physical property of the environmental object rather than with subjective aspects of my experience. The crucial move here is that if the color experience I have on viewing the rose is somehow defined by the physical properties of the rose, then the whole point of the target article is undermined. There is no subjectivity barrier to begin with, because color experiences are out in the world instead of inside my head. Therefore, everyone has access to the same (external) color experiences. My experience of the redness of the rose can't possibly be different from that of a pseudonormal individual because we are just encoding the same physical property (redness) by different neurophysiological mechanisms. Because the externalist objection is so potentially damaging to my views in the target article – and because I believe it to be so fundamentally mistaken – I will argue against it in some detail.

I believe that there are a number of serious problems with the externalist approach. First, it seems almost per-



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

verse to claim that redness is a coherent physical property in anything like the normal usage of this phrase in the physical sciences. The physical property that most strongly determines the perceived color of an object is its reflectance spectrum: the proportions of photons it reflects at each wavelength within the visible range (roughly 400–700 nm). The physical sense in which the rose itself can be said to be a particular shade of red (i.e., to produce in a normal human trichromat that particular experience of redness) is a very complex function of its reflectance spectrum, one that would be completely unmotivated except for its effect on the visual nervous system of the experiencing observer. A potentially infinite number of physically different reflectance spectra can give rise to the very same experience of redness, for example, because the infinite number of dimensions of reflectance spectra projects onto just three dimensions of human color experience. These exact perceptual matches despite physical differences are called "metamers," and their existence is a fundamental fact about color perception. About the only thing that metamers have in common is that they appear to produce identical experiences (and therefore no discrimination behavior) in particular organisms. As physical properties, then, perceived colors of objects do not seem to "carve nature at its joints" except with specific reference to color experience.

Even so, let's expand the usual notion of physical properties so that we can call the color a person experiences when viewing a given object one of that object's physical properties, simply because it can be determined, at least in principle, from its physical structure. Notice, however, that the same object will have to have different objective color properties for different equivalence classes of observers. The rose will not cause the same experiences in different types of color blind individuals and color weak individuals. Moreover, if one takes the broad view of what constitutes color experience (see sect. R2), the same rose will have a different physical color for *every* observer, and this objective color will have to include virtually any perceivable property.

A second problem is that, even if we restrict color experience to the narrow view, there are a lot of other conditions that must be imposed before we could predict the precise color the observer would experience on viewing the rose. We would need to know not only the rose's reflectance spectrum, but the reflectance spectra of surrounding surfaces, the illumination spectrum of the illuminating light source, and the pattern of shadows and reflected light produced by nearby surfaces in the environment. We would have to know these things because they all affect the precise color a person experiences when viewing the same object with the same purported physical color. Only if color constancy were perfect – and it is not, by a long shot – could one hope to equate the experience of redness with the reflectance properties of the rose itself.

Even so, all these other factors that must be known are (or could be formulated as) physical properties of physical objects as viewed in a particular context and from a particular position, so we could just fold some physical description of all of them into the externalist view of color. Other problems arise, however, when the state of the observer differs. The same rose will look less saturated after the observer has adapted to an intensely red region in the same area of the visual field, and it will look more saturated if the adapting region is its complementary shade of green.

Even so, these adaptational effects depend on the observer's viewing history with physical objects, so perhaps we can fold some physical description of all this recent viewing history into the externalist definition of color experience. However, there are color experiences that arise in the absence of *any* physical object, such as colored dreams and colored hallucinations. In the case of some dreams and certain kinds of hallucinations (like mirages), such color experiences might be based on specific memories of physical objects, in which case one could conceivably track down the objects in question and point to their physical properties in absentia. But in other cases of hallucinations, such as migraine headaches and even pressing on the eyeballs with one's fingers, there is no object to whose physical properties the color experiences can be attributed, and yet the person still experiences colors. The physical properties of an object are therefore not necessary for the experience of color. There are also clear cases of color illusions in which the object has the "wrong" physical properties to produce the experienced color. Chromatic afterimages and simultaneous color contrast are two well-known examples, and subjective colors induced by lines flickering at particular rates (as discussed in **Lockhead & Huettel**'s commentary) are another. It seems the only way to deal with these problems is to posit some nonexistent object or even the wrong object – the one that the subject appears to see – and say that their physical properties constitute the visual experience. But this seems nonsensical. What is present in each and every case is the internal experience, so why twist ourselves into knots to get some corresponding object with the corresponding physical properties into the theoretical picture?

A further complexity that is difficult to understand in externalist terms is that observers can be in different modes of perceiving color at different times (see Palmer 1999, p. 313–14). In "distal mode" perception, people's color experiences are geared toward color constancy, so that the color they perceive is mainly determined by the reflectance spectrum of the surfaces. This causes externalists no problem, of course, because it is this mode of perception that they explicitly use as the rationale for their externalist views. But in "proximal mode" perception, people's color experiences are more closely attuned to the appearance of regions in the image on the retina, which are not generally constant over the object's projected image, even if the object's surface reflectance is entirely uniform. Shadows cast by other objects, shading caused by gradients of surface orientation or illumination, and highlights reflected from specular surfaces are just three obvious cases in which differences occur in experienced color during proximal versus distal modes of perception. Clearly this would require a second externalist approach, one that designates the physical spectra of the light within different image regions as the external "object" (rather than the actual physical objects). Even adding this second kind of external entity leaves important things out, however. One is the fact that in either mode, the perception appears to be a blend of both stimulus conceptions, albeit with different weighting functions. Another is the importance of the internal variable that determines the mode in which the perceiver is currently experiencing.

All this argues against accepting an externalist view of color experience. An additional problem I have with the externalist argument **Ross** makes is that I don't believe it will work even in the case of pseudonormal vision, to which he


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

applies it. If one takes an externalist view, I can see the rationale for claiming that a normal and a pseudonormal observer both see the same object as a particular shade of red because it has, by definition, the same physical properties. But now consider within-subject cases. Suppose a person becomes pseudonormal after some effective physiological intervention. The internalist prediction is that, provided the change is "swift and enormous" enough and that one can rule out certain memory changes (see sect. R7), the subject will now see the same rose as green rather than red and be able to report the change. It seems that the externalist must predict that he or she will continue to see it as red because it is, after all, the same object with the same physical properties. To rule out memory artifacts, such as those suggested by **Dennett, Bradie,** and **Pauen,** the intervention could be carried out only for one half of the visual field. In this case internalists predict that the half of the rose will continue to look red, whereas the other half will now look green. A true externalist would predict no difference between the parts of the rose that fall in two different halves of the visual field. Obviously, I do not know what the results of such experiments would be, but the externalist predictions seem (to me) strikingly implausible.

**R4. The isomorphism constraint.** The wedge I used to separate the behaviorally unknowable intrinsic qualities of experience from their behaviorally knowable relational structure is the isomorphism constraint. This device too comes under fire from various commentators. **Harrison** challenges the coherence of the notion of isomorphism itself, at least for this particular enterprise, **Wright** questions certain aspects of the formulation I use to explain isomorphism, **O'Brien & Opie** raise the issue of how intrinsic versus extrinsic representations fit into the notion of isomorphism, and **Nida-Rümelin** presses for more relaxed criteria to encompass justified beliefs about color experience.

Although **Harrison** begins by implicitly endorsing isomorphism as at least superficially attractive – he once entertained the idea himself – he ultimately decides that it is incoherent. The problem he identifies, at least as I understand it, is that two qualitatively disparate sets of sensory experiences simply cannot exhibit the same relationships unless they are the same experiences. The argument is a subtle one that I believe I have also worried about, but I came to a different conclusion than he did for reasons that I will now explain.

The difficulty **Harrison** alludes to (or so I believe, for I am not sure I fully understand his point) is that two sets of sensory experiences cannot be isomorphic in the sense of exhibiting or standing in the *same relationships* to one another, unless they are, in fact, the very same sensory experiences. For example, I would claim that the lightness dimension of color experience could be isomorphic to another experiential dimension, such as pitch height in tonal experience or some alien dimension of visual experience in which lightness experiences are replaced by corresponding values on some other, completely different sensory dimension. Harrison would object that when the elements of color experiences are mapped systematically to experiences in another domain, the relationships among corresponding elements are actually *not* the same. I take this to mean that he believes that the relations among sensory experiences are as incommensurable as the sensory experiences themselves are. In what way can *lighter than* relations be the same as *higher pitch than* relations? If the relations are actually the same, then the experiential dimensions must also be the same.

It is for this reason that I was, or tried to be, careful to say that by "isomorphic" I meant that they have the same *relational structure* rather than if they have the same *relationships.* The distinction is important, because although the relations of "lighter than" between pairs of colors is indeed profoundly different from the relation of "higher than" among tones, there are ways of formalizing these relations so that they are the same at the abstract level of their structure (see sect. R1). Note 3 of the target article spells out the notion of "isomorphism" for one particular example in terms of Tarski's (1954) model theory in such a way that the sameness of their abstract relational structures is laid bare. I conclude that this notion of isomorphism, stripped of what **Harrison** calls the "primitive elements" (or "color-presentations") and of what I call "intrinsic qualities," demonstrates the coherence of the approach that he rejects as incoherent.

**Wright** comments on the relation of Sellars's (1922) early discussion of isomorphism to my own discussion. I was not aware of Sellars's treatment, which I thank Wright for calling to my attention. There is an important difference in our use of the concept of isomorphism, however, which I mentioned in Note 11 of the target article. Sellars appears to be talking about what I called "psychophysical isomorphism:" an isomorphism between environmental objects and internal representations. I am talking about isomorphism between one person's internal sensory representation and another person's. As I say in that note, I am skeptical about the general validity of the former, especially for color, but quite confident about that of the latter.

**Wright** then objects to my talking about isomorphism in terms of a mapping of objects that preserves relational structure. I did so because that is the way Tarski (1954) formalized the notion of isomorphism in his model theory, and that, as I explained in my replies to **Myin** and **Harrison,** is how I understand the notion of "same structure" abstracted from concrete relations. Wright's real point, however, appears to be that I did not worry about how the "objects" of experience are formulated in different observers. I was assuming that scientists were making such decisions in their theoretical analyses and that they would do so in uniform ways, but this is not necessarily the case, as he points out. He refers to this as a "social parameter," because there must be agreement on the nature of the "objects" to fix on an isomorphism. Indeed, differences between people in different cultures in terms of what they take to be the "objects" of their experience that are picked out by words may underlie at least some of the controversy over cross-cultural variations in color perception and naming (e.g., Saunders & Van Brakel 1997). **Frumkina** also expresses concerns over sociocultural influences on color experience, and **Malcolm** may be making a similar point in his commentary when he refers to the role of intersubjective agreement and social context in "conceptualizations" about color experience.

I really do not know quite what to make of or say about these objections based on social and/or cultural factors. Surely conceptualizations, social influences, and cultural variations will play some role in the scientific understanding of the nature of color experiences. My own belief is that they will turn out to be relatively minor compared with



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

more basic sensory and perceptual processes, but I have no knock-down argument that this is the case.

**O'Brien & Opie** take a different approach to the isomorphism constraint, one that is strongly connected to my own earlier distinction between intrinsic and extrinsic forms of representation (Palmer 1978). Although I have always liked this distinction and wish that it had received more attention than it did, I am not convinced that O'Brien & Opie's application of it is entirely justified.

Intrinsic versus extrinsic representations distinguish between representations that preserve relational structure of their source domain in critically different ways. Intrinsic representations do so by having the same inherent structure in their own corresponding relations as in the source relations they represent. Extrinsic representations do so merely by mimicking relational structure in a less constrained relation. The "structure" at issue refers to logical properties of (i.e., constraints on) represented and representing relations, such as transitivity and symmetry (see sect. R1). By "inherent" structure of a relation, I mean relational constraints that cannot be broken. As I mentioned above, the relation "lighter than," for example, is inherently transitive (if A is lighter than B and B is lighter than C, then A is lighter than C) and antisymmetric (if A is lighter than B, then B cannot be lighter than A). Color spaces represent this relation intrinsically by the "higher than" relation in space, which is likewise inherently transitive and antisymmetric. "Lighter than" could be represented extrinsically by the relation "points to" in a network of nodes that represent color experiences and directional arrows that represent the "lighter than" relation. "Points to" in a directional network is neither inherently transitive (if A points to B and B points to C, then A might point to C or it might not) nor inherently antisymmetric (if A points to B, then B might point to A or it might not). "Points to" can still represent "lighter than," however, because it is less constrained than "lighter than," thus allowing it to take on the required structure simply by reflecting the empirical constraints in the source domain. This is the essence of the intrinsic/extrinsic distinction.

I believe that **O'Brien & Opie** are wrong in asserting that functionalism is committed to extrinsic representation. Functionalism is indeed committed to causal relational structure, as they say, but causal relational structure can be embedded in either intrinsic or extrinsic representing relations. They then contrast their view of functionalism with what they call "structuralism," which appears to be defined by a commitment to intrinsic representation and lies, according to them, midway between functionalism and the biological conception. They further suggest that conscious experience can be identified with "structural roles" that appear to be defined by physical relations rather than causal ones. In my understanding of their view, structuralism would just be a subset of standard functionalism that insists on intrinsic representations. Moreover, I do not see any clear connection between intrinsic representations and conscious experience, especially given that the color machine (Fig. 6 in the target article) uses intrinsic representations of color relations, yet is unlikely to produce color experiences.

Nevertheless, **O'Brien & Opie**'s comments raise the interesting issue of how the intrinsic/extrinsic representational distinction relates to the ideas expressed in the target article, and therein lies a puzzle. On the one hand, the intrinsic/extrinsic distinction appears to reside at the subisomorphic level. The rationale for this claim is that if two representations (one intrinsic and one extrinsic) preserve the same relational structure about the same source domain, then they are necessarily isomorphic to each other and cannot be distinguished strictly on the basis of representational aspects. If the difference does lie at the subisomorphic level, then I should further have to agree that intrinsic and extrinsic representations cannot be discriminated by behavioral means. But I originally formulated the distinction to clarify the analog/propositional debate, proposing that analog representations were intrinsic and propositional representations were extrinsic, and there is a mountain of behavioral evidence that has been interpreted as discriminating between analog and propositional representations. So there is a conceptual muddle in here somewhere. I would guess that it results from differences in how intrinsic and extrinsic representations are *processed.* That is, the distinction might actually be below the level of isomorphism in terms of representations but above it in terms of processes.

**Nida-Rümelin** remarks on a very different aspect of the isomorphism constraint. She suggests that although the isomorphism constraint is true if it is interpreted in terms of what can be known with scientific certainty about intrinsic qualities of experience or what can be detected by third-person observers about these qualities, it is false if it is interpreted in terms of "justified beliefs" about them. The issue, of course, is what standard of evidence one takes for "justified beliefs." Obviously, the bar is lower for justified beliefs than for scientific certainty, but how much lower?

In my target article, I was talking about scientific standards of evidence for drawing scientific conclusions, not simply justified beliefs. Although scientific conclusions are not truly certain in the same way that logical proofs are, scientific standards are at least much higher than everyday standards for everyday beliefs. **Nida-Rümelin** has formulated a set of conditions that she offers for justified beliefs about color experiences, and although I do not pretend to have a strong view about what constitutes a justified belief, I have no problem with her suggestions. My only concern is that by sneaking in the qualifier "scientifically based" before "justified phenomenal belief," she might inadvertently give some scientific credibility to beliefs justified in this way. I prefer to interpret this qualifier literally, as indicating merely that the beliefs in question are based on scientific evidence about the brain, not that any conclusions reached using these relaxed criteria are endorsed by science.

**R5. The color machine, the color room, and functionalism.** Several of the commentators zeroed in on the color machine as a weak argument against functionalism, mainly because my "one trick pony" was overly simple (e.g., **Hardin, Saunders, Van Gulick,** and **Viger**). My stated intention was that the machine diagrammed in Figure 6 was just the "front end" of a larger system that would do all the standard color tasks within the "narrow view" of color experience (see sect. R2). I did not envision that it would have to say which colors it liked better, whether red or blue made it feel more relaxed, or even what colors ripe versus unripe tomatoes were. The objections these commentators raised to this machine were (1) that there is a lot more to color experience than these basic tasks imply (see sect. R2) and (2) that to produce experience, the entire system is required,


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

not just some color-relevant piece of it. Although I admit that both of these statements are plausibly true, they are not sufficient to convince me that functionalism contains the kind of structure required to explain color experience.

The prototypical claim of the commentators is that if the color machine were made sufficiently complex (presumably in the right way), it would *have* color experiences rather than merely simulating them. My problem with this sort of complexity argument is that it is unclear to me why or how experiential qualities would arise from positing more complex computations. The situation reminds me of the cartoon depicting a professor resembling Einstein doing a mathematical proof at the blackboard up to some point, at which he writes, "Then a miracle happens!" and finishes the proof. If experience exists because of additional complexity per se, then somewhere amidst the additional computations a miracle must indeed happen. And why should it happen just by connecting the internal color representation to other sorts of color-related representations, such as preferences, mood associations, and characteristic colors of objects? I doubt that the miracle is just more computations of the standard sort – even many, many more – and functionalists saying that it is doesn't make it so.

Neither does my doubting it make it false, of course, for my skepticism about this holist/functionalist complexity response is grounded in intuitions that may turn out to be wrong. I doubt, for example, that a computer can ever be programmed to have color experiences, no matter how complex its computations might be, how sophisticated its language capabilities might be, or what type of simulated "self-reflection" on its own internal representations is built into it. On the contrary, I also believe that there are probably a large number of biological creatures that do have color experiences of some sort despite having pretty simple brains, no language at all, and probably no self-reflective abilities at all. This leads me to believe that functional/computational complexity is unlikely to be the magic ingredient, but I do not claim to have proved this by revealing my biases.

Other commentators view the color room argument with similar skepticism (e.g., **Byrne, Jakab, Schröder, Van Gulick, Viger**). Several claim that Searle's original Chinese room argument has been refuted, either by the "systems reply" or by some other argument made since Searle's article appeared. I will try to side-step this hornet's nest of problems by stating that *if* Searle's Chinese room argument has been decisively refuted, then so has the color room argument, because they are really the same argument. I am not convinced that it has been, but others obviously are. In any case, I will not attempt to debate that issue here.

Another front on which the color room has been attacked is a version of the complexity argument. I may not experience color when I am computing in the color room, but then neither does the color-processing portion of the brain necessarily experience color in isolation from the rest of the brain (e.g., **Jakab**). I suppose the only reply to this objection is to presume that I am in the "mind room" and have the task of computing everything, color included. The problem is that this move eliminates a good deal of the appeal of the color room, which was the relative simplicity of the computations and the strength of the intuition that it would not have color experiences.

**Jakab** also sketches an alternative functional analysis of experience as arising when a subject (one's self) *entertains* a pattern of activity in the appropriate area of the brain; for example, color arises from the right sort of firings taking place in V4 neurons. The color room argument then does not work because the computations aren't occurring in the right way or even in the right part of the brain. There may indeed be something to this idea, but it doesn't sound like functionalism to me, primarily because it makes such heavy reference to brain mechanisms. One of the main tenets of functionalism is the notion of multiple realizability – the hypothesis that the same mental state can be achieved by equivalent computations in different physical devices. Jakab's formulation thus runs perilously close to contradicting a central tenet of functionalism.

It is comforting to know that at least one reader found the arguments against functionalism convincing. After endorsing the conclusion that functionalism is inadequate to the task of understanding color experiences, **Howard** then goes on to ask what view should replace it as an adequate philosophical framework for color experience. His rather surprising answer is eliminative materialism (e.g., Churchland 1981). He justifies the choice by pointing to a close fit between Churchland's (1995) vector space formulation of neural networks and the framework of isomorphism in the target article. The problem is that color experiences seem very unlikely to be "eliminated" by descriptions in terms of neural activation patterns rather than simply "reduced" to neural events. The close correspondence that is already believed to hold between color perceptions and simple neural firing rates undercuts the possibility that there is any good reason to get rid of color experiences as theoretical entities in favor of the isomorphic activation patterns that underlie them. Eliminative materialism seems appropriate only when the mental entities in question are seriously flawed in relation to the more accurate scientific description of the underlying brain events. This does not appear to be the case with perceived colors, however.

**R6. Introspection versus behavior as scientific methods.** Both **Smythies** and **MacLennan** argue that one need not (or should not) rely on strict behavioral methods to understand color scientifically because we can rely on introspection instead. I do not deny that introspection almost always provides a crucial starting point in color science or any other psychological science. I do object to the stronger claim that it is sufficient, either alone (Smythies) or in conjunction with neurological theories (MacLennan) for doing science. One problem is that unless one is willing to embrace the idea that investigations that are limited to one's self constitute science – that is, that all one needs to do is convince oneself – behavior of some sort must be included, and the subjectivity barrier raises its ugly head. Even describing one's introspections (as Goethe did according to MacLennan's example) verbally to someone else involves behavior. Because introspections are inherently subjective, they cannot be the basis of objective generalizations (which I take to be necessary for any scientific inquiry) unless they are externalized by some sort of behavior.

Furthermore, when two people's introspective analyses differ, there is no way to decide which is correct without behavioral measures. One cannot turn to neurological theories or data as the deciding evidence, as **MacLennan** suggests, because the disputed fact concerns the nature of the experiences themselves, not whatever neural events might underlie them. Having the wrong neural theory could then



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

lead one to reject an accurate phenomenological analysis that was inconsistent with it. This possibility may seem farfetched, but something like it almost happened when proponents of Helmholtzian trichromatic color theory attempted to reject Hering's analysis of color into four chromatic primitives (red, green, blue, and yellow) because it conflicted with the three hypothetical receptor types in Helmholtz's neural theory.

**R7. Empirical issues about color experience.** Many of the commentators made what I consider to be straightforward empirical points concerning scientific facts about color experience, color naming, or related topics. The main generalization is that color experience is much more complex than I indicated in my bare-bones description of basic phenomena of color vision. With this I generally agree. Below I briefly state some of the most important clarifications and explain, where I can, why I did not discuss them. I also indicate points with which I do not agree and explain why I do not.

**Hardin, van Brakel,** and **Saunders** all mention in one way or another that there are actually more color spaces (e.g., Munsell, CIE, NCS) and color models (e.g., Land's retinex theory) rather than just the view I present. This is true. Because these other color structures are constructed according to different psychometric and scaling principles, they reflect different relational properties among color experiences and have somewhat different symmetries – if indeed they have any at all. **MacLaury** and **Kay** similarly mention metric properties that break symmetries, such as the fact that there are more just noticeable differences (jnd's) between red and blue than between blue and green. **Griffin** actually presents what I take to be new data on similarity judgments that rule out the existence of any precise symmetries, although he makes the sensible point that there can be degrees of symmetry, with gradations of symmetry-breaking.

The primary implication of these (and other) facts is that, on closer inspection, color experience does not have the three symmetries I suggested in Figure 3 of the target article. Such empirical constraints further undermine the possibility that the color experiences of two individuals could be the same, yet differently connected to the outside world. This conclusion holds, however, only if one assumes automorphism, that is, that all normal trichromats have the same color experiences.

I intentionally avoided these kinds of complications in my original target article for four reasons. First, my goal was to elucidate the nature of the argument concerning how empirical constraints bear on the basic question of detecting color transformations rather than to arrive at a definitive answer to Locke's thesis. Second, I wanted to base my analysis as much as possible on robust qualitative phenomena rather than on more questionable quantitative relations that depend importantly on the details of specific psychophysical methods. For example, is the number of jnd's between two colors the appropriate way to measure the metric aspects of color experience, or should one use more direct ratio scaling procedures? Third, as **Block** mentions, it can be argued that the mere possibility of there being symmetries in color experience is sufficient to reject functionalism, even if human color vision contains none. Fourth, and most important, I later use the possibility of alien colors and color zombies to argue that the automorphism assumption is unnecessary to the general question, thus rendering irrelevant the issue of whether or not the structure of one's color experience contains any symmetries.

There is one empirically based objection to my line of argument with which I must take exception, however. If I understand his argument, **Cohen** claims that the set of color transformations I suggest might be undetectable is incomplete because it does not include slight rotations, stretches or squeezes.[3] The problem is that such transformations would, in fact, be behaviorally detectable in appropriate quantitative psychophysical tasks. In the terminology of my target article, individuals whose judgments of, say, unique hues differ significantly are members of different behavioral equivalence classes within the larger class of so-called "normal trichromats." In fact, there are a number of well known kinds of "color weakness" that I did not discuss that result from the presence of slightly different photopigments in retinal cones. I therefore do not believe that Cohen's objection is a problem with the analysis I gave.

Other writers comment on various empirical matters concerning BCTs. **Kay** presents some intriguing data about complexities in the relation among derived and composite BCTs, and **Paramei** argues that *goluboi* is not a BCT of Russian. I have no quarrel with these presentations, but see them as only tangentially relevant to the main points of the target article. Specifically, I remain unconvinced that BCTs are properly considered part of one's "basic" color experience on viewing the world, but, if they are, then the issues they raise are relevant to answering Locke's question about inverted spectra.

Still other commentators discuss aspects of color vision that I did not consider in the target article, but should have. **Lockhead & Heuttel** remind us of the mysterious "subjective" color experiences induced by particular kinds of moving or flickering line displays. Nobody knows why these stimuli should cause such experiences, but they are certainly relevant to the issue of whether it makes sense to define color as an internal or external phenomenon (see sect. R3). **Humphreys & Riddoch** and **Benson** report on the importance of neuropsychological studies of patients with various deficits in color perception. The findings they discuss, and others like them, bear importantly on the viability of within-subjects designs for detecting changes in color experience (see sect. R8) and on the possibility of color zombies.

**Kranda** and **Backhaus** discuss the target article with respect to the genetics, evolution, and physiological structure of color vision systems. Kranda argues that the probability of switching the long- and medium-wavelength cones would be vanishingly small ($10^{-12}$), and evolutionarily improbable as well.[4] Backhaus presents his neuronal color opponent coding theory in the honeybee and argues that the kind of physiological changes required for color transformations would be highly unlikely because several mutations would be required simultaneously.

I am not knowledgeable enough in these fields to remark on the technical arguments that these authors present. By way of a general response, however, I would like to point out that (1) logical possibility is not the same as low probability and my arguments were about logical possibilities, (2) genetic probabilities under natural selection may be less relevant in the not too distant future with the advent of genetic engineering, and (3) none of their arguments address


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

either the alien color argument or the color zombie argument, both of which I take to be serious considerations in understanding color experiences. I also believe that **Backhaus** is overly optimistic when he claims that the holy grail of physics, a grand unified "theory of everything," must necessarily allow us to compare color experiences in different brains. That sort of naive acceptance of the reach of physical science is exactly what the target article was questioning, and I still see no good reason to believe that Backhaus will turn out to be correct in his assessment.

**R8. The problem of memory in within-subject experiential changes.** Some of my arguments about the power and usefulness of within-subject designs rest on intuitions about what would happen as the result of certain biological interventions. If the photopigments in medium- and long-wavelength cones could magically be interchanged, for example, I assumed that the subject would notice the change and be able to report it behaviorally to an experimenter, provided the changes were "swift and enormous."

Several commentators (e.g., **Dennett, Bradie,** and **Pauen**) took exception to this assumption for various reasons. Bradie claimed that for the purposes of qualia comparison, one's prior self is as alien a being to one's present self as is any other mind. Dennett claimed that we could all gradually become color zombies and not know it. Pauen suggested that experiential changes would be undetectable if my memories changed in the same way as my experience did.

As convenient as these possibilities might be from certain theoretical points of view, the facts indicate that they are not serious concerns. There are clear cases, cited in commentaries by **Humphreys & Riddoch** and by **Benson,** in which neurological patients clearly detect changes in their color experiences and report them to their physicians. Such effects are not limited to neuropsychological studies, however. Putting on tinted goggles or (for some people, at least) taking a large enough dose of Viagra causes a swift and/or enormous enough effect to notice a bluing or greening change in the appearance of objects. The memories of these people do not change concomitantly, and their present selves' access to the color experiences of their prior selves is manifestly sufficient to detect the alterations in color appearances.

The fact that some people under some conditions are able to detect color differences does not entirely answer the critics' concerns because there might be other people under other conditions in which memory changes would mask experiential changes. We therefore would like to devise a form of physiological intervention in which memory plays little or no role in people's ability to detect any color changes it might produce. Perhaps the best way to accomplish this is to envision carrying out the physiological intervention so that it affects only half of the visual field. Cortical interventions would then be performed in only one cerebral hemisphere, and lower level interventions to the appropriate portion of the retina or LGN.

The observer could then view half of an object or scene with the unaffected hemifield and the other half with the affected hemifield and compare the two simultaneously, with no memory component to distort the results. There are no such results that I know of, but we are talking about hypothetical (i.e., thought) experiments to begin with, so this minor twist causes no real conceptual difficulty. I therefore do not feel compelled by the memory objection to change my basic claims about the importance of within-subject designs.

The commentaries on the target article have challenged, but not changed, my belief that color is one of the most interesting and informative areas of cognitive science, and one that can usefully serve as a model for investigating other areas, including their experiential versus behavioral aspects. I hope that the target article, commentaries, and reply have advanced our understanding of at least the questions underlying color experience, if not the answers.

NOTES

**1. Dennett** makes an odd and misleading analogy that access to the intrinsic qualities of one's sensory experience is like access to whether one's left or right hemisphere is dominant. The important difference is that nobody claims to have introspective access to which of one's hemispheres is dominant, whereas everybody (except perhaps Dennett) claims to have introspective access to the intrinsic qualities of their own sensory experiences.

**2.** This scheme would not work if the wiring from M- and L-cones to the R-cell and the G-cell were determined by the nature of the pigment, but I know of no logical reason why this needs to be so.

**3. Cohen** also claims that I was trying to save functionalism – despite the fact that I ended up arguing against it – but I was not. I only wanted to give functionalism its best shot.

**4. Kranda** also states that the H, B, and S outputs of the color machine in Figure 6 would always equal zero, but this is untrue. I intentionally did not specify the weightings on the connections to the hypothetical H, B, and S nodes because the nature of the computation has not been discovered and is not relevant to the argument I was making. The arrows to the output nodes in the diagram merely indicate which opponent-process nodes presumably affect which output nodes, not how they interact.

## References

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


Abramov, I., Gordon, J., Akilov, V., Babiy, M., Bakis, G., Ilyusha, S., Khamermesh, K. & Vayner, A. (1997) Color appearance: Singing the Russian blues. *Investigative Ophthalmology and Visual Science* 38:S899. [GVP]

Armstrong, D. M. (1968) *A materialist theory of the mind.* Routledge. [AB]

Baars, B. (1988) *A cognitive theory of consciousness.* Cambridge University Press. [aSEP]

Bachelard, G. (1972) *Le matérialisme rationnel.* Presses Universitaires de France. [BH]

Backhaus, W. (1992) Evolutionary aspects of the theory of color vision in insects. In: *Proceedings of the 3rd International Congress of Neuroethology, McGill University, Montreal, Quebec, Canada, August 9–14, 1992.* [WB]

(1993) Color vision and color choice behavior of the honey bee. In: *Recent progress in neurobiology of the honey bee.* Special issue. *Apidologie* 24:309–31. [WB]

(1994) Color sensations in honeybees? In: *Society of Neuroscience Abstracts, vol. 20, part 2: Visual psychophysics and behavior,* 647.14. [WB]

(1995) On the proof of color sensations in animals. Eighteenth European Conference of Visual Perception, Tübingen, Germany. Abstracts. *Perception* 24:91a, Supplement. [WB]

(1996) Allgemeine Sinnesphysiologie. In: *Neurowissenschaft. Vom Molekül zur Kognition,* ed. H. J. Dudel, R. Munzel & R. F. Schmidt. Springer. [WB]

(1997a) Colour sensations in honeybees? In: *John Dalton's colour vision legacy: Selected Proceedings of the International Conference,* ed. C. Dickinson, I. Murray & D. Carden. Taylor & Francis. [WB]

(1997b) On the problem of physiological modeling of color sensations. In: *From membrane to mind. Proceedings of the 25th Göttingen Neurobiology Conference, vol. II,* N. Elsner & H. Wässle. Thieme. [WB]

(1997c) On the constraints for physiological models of color sensations. (Über die Randbedingungen für physiologische Modelle von Farbempfindungen). In: *Verhandlungen der Deutschen Zoologischen Gesellschaft,* ed. D. Zissler. Fischer. [WB]




https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

(1998a) Physiological and psychophysical simulations of color vision in humans and animals. In: *Color vision - perspectives from different disciplines,* ed. W. Backhaus, R. Kliegl & J. S. Werner. De Gruyter. [WB]
(1998b) On the constraints for a physiological model of color sensations. In: *New neuroethology on the move. Göttingen Neurobiology Report 1998: Proceedings of the 26th Göttingen Neurobiology Conference 1998, vol. I,* ed. N. Elsner & R. Wehner. Thieme. [WB]
(1998c) Neuronal color coding in the honeybee. In: *From structure to information in sensory systems,* ed. C. Taddei-Ferretti & C. Musio. World Scientific. [WB]
(1998d) The internal representation of color information in humans and animals. In: *Downward processes in the perception representation mechanisms,* ed. C. Taddei- Ferretti & C. Musio. World Scientific. [WB]
(1998e) Conscious and unconscious color vision in man and animals. In: *Downward processes in the perception representation mechanisms,* ed. C. Taddei-Ferretti & C. Musio. World Scientific. [WB]
(1999) Neuronal coding and color sensations. In: *Proceedings of the 5th International Work-Conference on Artificial and Neural Networks (IWANN '99), Alicante, Spain, June 2–4, 1999: Vol. I. Foundations and tools for neuronal modeling,* ed. J. Mira & J. V. Sánchez-Andrés. Springer. [WB]
(in press) Physiological modeling of color sensations. In: *Neuronal bases and psychological aspects of consciousness,* ed. C. Taddei-Ferretti & C. Musio. World Scientific. [WB]
Backhaus, W., ed. (in preparation) Neuronal coding of perceptual systems. Proceedings of the International School of Biophysics, Course on "Neuronal coding of perceptual systems," Ischia/Naples, Italy, October 12–17, 1998. Istituto Italiano per gli Studi Filosofici, Study Program on "From neuronal coding to consciousness." In: *Series of biophysics and biocybernetics, vol. 8,* ed. W. Backhaus. World Scientific. [WB]
Backhaus, W., Gerster, U., Buckow, H., Pielot, R., Breyer, J. & Becker, K. (1996) Physiological simulations of neuronal color coding in honeybees. In: *Bionet '96. Bio-informatics and pulspropagating networks - selected contributions. 3rd Workshop, Berlin, November 14–15, 1996.* GFAI. [WB]
Backhaus, W., Kliegl, R. & Werner, J. S., eds. (1998) *Color vision - perspectives from different disciplines.* de Gruyter. [WB]
Backhaus, W. & Kratzsch, D. (1993) Unique-colors in color vision of the honeybee? In: *Genes, brain and behavior. Proceedings of the 21st Göttingen Neurobiology Conference,* ed. N. Elsner & D. W. Richter. Thieme. [WB]
Becker, K. & Backhaus, W. (1998) Physiological modeling of temporal properties of the neuronal color coding system in the honeybee. In: *From structure to information in sensory systems,* ed. C. Taddei-Ferretti & C. Musio. World Scientific. [WB]
(in press) A quantitative model describing the electrophysiological properties of dark and light adapted photoreceptors in the honeybee (*Apis mellifera)* worker. *Biological Cybernetics.* [WB]
Berlin, B. & Kay, P. (1969) *Basic color terms: Their universality and evolution.* University of California Press. [PK, BM, aSEP, GVP]
Block, N. (1978) Troubles with functionalism. In: *Minnesota studies in the philosophy of science, vol. IX,* ed. C. W. Savage. (Reprinted in: *Mind and cognition,* ed. W. Lycan. Blackwell, 1990; *The nature of mind,* ed. D. M. Rosenthal. Oxford University Press, 1991; *Philosophy of mind,* ed. B. Beakley & P. Ludlow. Mit Press, 1992; *Readings in philosophy and cognitive science,* ed. A. Goldman. MIT Press, 1993.) [NB]
(1980) What is functionalism? In: *Readings in philosophy of psychology, vol. 1,* ed. N. Block. Harvard University Press. [NB]
(1990) Inverted earth. *Philosophical Perspectives* 4:51–79. [NB]
Reprinted in: *The nature of consciousness,* ed. N. Block, O. Flanagan & G. Güzeldere. MIT Press, 1997. [AB]
(1995) On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18:227–47. [PJB]
Reprinted with changes in: *The nature of consciousness,* ed. N. Block, O. Flanagan & G. Güzeldere. MIT Press, 1997. [AB]
(1999) Sexism, racism, ageism and the nature of consciousness. In: *The philosophy of Sydney Shoemaker,* ed. R. Moran, J. Whiting & A. Sidelle. *Philosophical Topics* 26(1&2). [NB]
Block, N., ed. (1980) *Readings in philosophy of psychology, vol. 1 and 2.* Harvard University Press. [JvB]
Block, N., Flanagan, O. & Guzeldere, G. (1997) *The nature of consciousness: Philosophical debates.* MIT Press. [NB]
Block, N. & Fodor, J. (1972) What psychological states are not. *Philosophical Review* 81:159–81. [NB]
Boghossian, P. & Velleman, D. (1989) Color as a secondary quality. *Mind* 98:81–103. (Reprinted in: *Readings on color: The philosophy of color,* ed. A. Byrne & D. Hilbert. MIT Press). [NB]
(1991) Physicalist theories of color. *Philosophical Review* 100:67–106. (Reprinted in: *Readings on color: The philosophy of color,* ed. A. Byrne & D. Hilbert. MIT Press). [NB]
Bornstein, M. H., Kessen, W. & Weiskopf, S. (1976) Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance* 2(1):115–29. [aSEP]
Boster, J. (1986) Can individuals recapitulate the evolutionary development of color lexicons? *Ethnology* 25:61–64. [PK]
Boynton, R. M. (1979) *Human color vision.* Holt, Rinehart & Winston. [MN-R, PWR]
Boynton, R. M., Schafer, W. & Neun, M. E. (1964) Hue-wavelength relation measured by color-naming method for three retinal locations. *Science* 146:666–68. [GVP]
Bulmer, R. H. N. (1968) Karam colour categories. *Kivung* 1:120–33. [JvB]
Byrne, A. & Hilbert, D. (1997) Colors and reflectances. In: *Readings on color, vol. 1: The philosophy of color,* ed. A. Byrne & D. Hilbert. MIT Press. [NB, AB]
Cataliotti, J. & Gilchrist, A. (1995) Local and global processes in surface lightness perception. *Perception and Psychophysics* 57(2):125–35. [rSEP]
Chalmers, D. J. (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2:200–19. [BM]
(1996) *The conscious mind.* Oxford University Press. [BM, JTT]
Chapanis, A. (1965) Color names for color space. *American Scientist* 53:327–46. [RMF]
Chisholm, R. (1954) Sellars' critical realism. *Philosophy and Phenomenological Research* 15(1):33–47. [EW]
Church, J. (1998) Two sorts of consciousness? *Communication and Cognition* 31:57–72. [EM]
Churchland, P. M. (1981) Eliminative materialism and the propositional attitudes. *Journal of Philosophy* 78:67–89. [HH, rSEP]
(1986) Some reductive strategies in cognitive neuroscience. *Mind* 95:279–309. [HH]
(1989) *A neurocomputational perspective: The nature of mind and the structure of science.* MIT Press. [HH]
(1995) *The engine of reason, the seat of the soul: A philosophical journey into the brain.* MIT Press. [HH, rSEP]
(1998) Conceptual similarity across sensory and neural diversity: The Fodor/Lepore challenge answered. *The Journal of Philosophy* 95(1):5–32. [GJO]
Clark, A. (1993) *Sensory qualities.* Oxford University Press. [AB]
Cohen, J. & Gordon, D. A. (1949) Prevost-Fechner-Benham subjective colors. *Psychological Bulletin* 46:97–136. [GRL]
Copeland, B. J. (1993) The curious case of the Chinese gym. *Synthese* 95:173–86. [AB]
Corbett, G. & Davis, R. (1997) Establishing basic color terms: Measures and techniques. In: *Color categories in thought and language,* ed. C. L. Hardin & L. Maffi. Cambridge University Press. [RMF]
Corbett, G. & Morgan, G. (1988) Colour terms in Russian: Reflections of typological constraints in a single language. *Journal of Linguistics* 24:31–64. [GVP]
Courtney, S. M., Finkel, L. H. & Buchsbaum, G. (1995a) Network simulations of retinal and cortical contributions to color constancy. *Vision Research* 33:413–34. [HH]
(1995b) A multi-stage neural network for color constancy and color induction. *IEEE Transactions on Neural Networks* 6:972–85. [HH]
Crane, T. (1992) The non-conceptual content of experience. In: *The contents of experience: Essays on perception,* ed. T. Crane. Cambridge University Press. [NB]
Crick, F. H. C. (1994) *The astonishing hypothesis: The scientific search for the soul.* Scribner. [aSEP]
Crick, F. H. C. & Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375:121–23. [aSEP]
Cross, S. A. (1994) Pathophysiology of pain. In: *Mayo Clinic Proceedings* 69:375–83. [MP]
Damasio, A. R. (1994) *Descartes' error: Emotion, reason, and the human brain.* G. P. Putnam. [EM, MP]
Davies, I. R. L., Corbett, G. G., Laws, G., McGurk, H., Moss, A. E. St. G. & Smith, M. W. (1991) Linguistic basicness and colour information processing. *International Journal of Psychology* 26:311–27. [GVP]
Dennett, D. C. (1978) *Brainstorms.* MIT Press. [aSEP]
(1991) *Consciousness explained.* Little, Brown. [DCD, aSEP, CDV]
De Valois, R. L. & De Valois, K. K. (1993) A multi-stage color model. *Vision Research* 33:1053–65. [BM]
De Valois, R. L. & Jacobs, G. H. (1968) Primate color vision. *Science* 162(3853):533–40. [aSEP]
Dretske, F. (1995) *Naturalizing the mind.* MIT Press. [NB, AB, PWR]
Dubois, D., ed. (1997) *Categorisation et cognition: De la perception au discours.* Kime. [RMF]
Dubois, D., Rouby, C. & Chevalier, G. (1997) Categorization for odors: What makes smell similar? In: *Proceedings of SimCat 1997: An interdisciplinary workshop on similarity and cognition.* University of Edinburgh Press. [RMF]
Fenton, M. (1997) Psychological approaches to the investigation of colour: Meaning and representation. *Spectrum* 11:4–9. [GVP]


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

Fodor, J. (1968) *Psychological explanation.* Random House. [aSEP]
(1981) The present status of the innateness controversy. In: J. Fodor, *Representations: Philosophical essays on the foundations of cognitive science.* MIT Press. [BS]
Fodor, J. & Lepore, E. (1992) *Holism: A shopper's guide.* Blackwell. [HH]
Frege, G. (1884/1950) *The foundations of arithmetic,* trans. J. L. Austin. Blackwell. [AB]
(1918/1988) Thoughts. In: *Propositions and attitudes,* ed. N. Salmon & S. Soames. Oxford University Press. (Reprint.) [AB]
Frijda, N. H. (1993) Moods, emotion episodes, and emotions. In: *Handbook of emotions,* ed. M. Lewis & J. M. Haviland. Guilford Press. [MP]
Frumkina, R. M. (1984) *Zvet, Smysl, Skhodstvo: Aspekty Psikholingvisticheskogo analiza.* Nauka. [RMF]
Frumkina, R. M. & Mikhejev, A. V. (1996) *Meaning and categorization.* Nova Science. [RMF]
Gage, J. (1993) *Colour and culture: Practice and meaning from antiquity to abstraction.* Thames & Hudson. [JvB]
Garner, W. R. (1974) *The processing of information and structure.* Erlbaum. [rSEP]
Gilchrist, A. L. (1988) Lightness contrast and failures of constancy: A common explanation. *Perception and Psychophysics* 43(5):415–24. [rSEP]
Goethe, J. W. von (1840) *Goethe's theory of colours,* trans. C. L. Eastlake. John Murray. [BM]
Goodwin, C. (1997) The blackness of black: Color categories as situated practice. In: *Discourse, tools, and reasoning: Essays on situated cognition,* ed. B. Lauren, L. Resnick, R. Säljö, C. Pontecorvo & B. Burge. Springer. [RMF]
Gordon, J., Abramov, I. & Chan, H. (1994) Describing color appearance: Hue and saturation scaling. *Perception and Psychophysics* 56:27–41. [GVP]
Gould, S. J. (1980) Is a new and general theory of evolution emerging? *Paleobiology* 6:119–30. [KK]
Gregory, R. L. (1981) *Mind in science.* Penguin. [JSm]
(1983) Perception: Philosophical issues. In: *The encyclopedic dictionary of psychology,* ed. R. Harré & R. Lamb. Blackwell. [EW]
Guo, K., Benson, P. J. & Blakemore, C. (1998) Residual motion discrimination using colour information without primary visual cortex. *NeuroReport* 9:2103–107. [PJB]
Hardin, C. L. (1988) *Color for philosophers: Unweaving the rainbow.* Hackett. [EM, JvB]
(1990) Color and illusion. Presented at The phenomenal mind: How is it possible and why is it necessary? Conference held by Zentrum fur Interdisziplinare Forschung, Bielefeld, Germany, May 1990. [aSEP]
(1997) Reinverting the spectrum. In: *Readings on color: Vol. 1. The philosophy of color,* ed. A. Byrne & D. R. Hilbert. MIT Press. [aSEP]
Harman, G. (1990) The intrinsic quality of experience. In: *The nature of consciousness,* ed. N. Block, O. Flanagan & G. Güzeldere. MIT Press, 1997. [AB]
(1996) Explaining objective color in terms of subjective reactions. In: *Philosophical issues 7: Perception,* ed. E. Villanueva. Ridgeview. [NB]
Harnad, S. (1987) Category induction and representation. In: *Categorical perception: The groundwork of cognition,* ed. S. Harnad. Cambridge University Press. [EW]
Harrison, B. (1967) On describing colours. *Inquiry* 10(1):38–52. [BH]
(1973) *Form and content.* Blackwell. [BH, aSEP]
(1986) Identity, predication and colour. *American Philosophical Quarterly* 23(1):105- 14. [BH]
(1987) Identity, predication and colour. In: *Philosophy and the visual arts,* ed. A. Harrison. Reidel. [BH]
Hayek, F. A. (1952) *The sensory order.* Routledge & Kegan Paul. [aSEP]
Heider, E. R. (1972) Universals in color naming and memory. *Journal of Experimental Psychology* 93(1):10–20. [aSEP]
Heisenberg, W. (1958) *Physics and philosophy.* Harper. [JSm]
Hering, E. (1878/1964) *Outlines of a theory of the light sense,* trans. L. M. Hurvich & D. J. Jameson. Harvard University Press. [BM, aSEP]
Heywood, C. A., Kentridge, R. W. & Cowey, C. (1998) Cortical color blindness is not "blindsight for color." *Consciousness and Cognition* 7:410–23. [PJB]
Hilbert, D. R. & Kalderon, M. (in press) Color and the inverted spectrum. *Vancouver Studies in Cognitive Science.* [AB]
Hoffman, D. (1998) *Visual intelligence: How we create what we see.* W. W. Norton. [CDV]
Hooker, C. A. (1995) *Reason, regulation and realism: Toward a regulatory systems theory of evolutionary epistemology.* State University of New York Press. [EW]
Humphreys, G. W., Troscianko, T., Riddoch, M. J., Bouvcart, M., Donnelly, N. & Harding, G. F. A. (1992) Covert processing in different visual recognition systems. In: *The neuropsychology of consciousness,* ed. A. D. Milner & M. Rugg. Academic Press. [GWH]
Hurvich, L. M. (1981) *Color vision.* Sinauer. [JC]
Indow, T. (1988) Multidimensional studies of Munsell color solid. *Psychological Review* 95:456–70. [GVP, JvB]
Jackson, F. (1982) Ephiphenomenal qualia. *Philosophical Quarterly* 32:127–36. [HH, MN-R]
James, W. (11890/1950) *Principles of psychology.* Holt. [aSEP]
Johnson-Laird, P. N. (1983) A computational analysis of consciousness. *Cognition and Brain Theory* 6:499–508. [aSEP]
Kandel, E., Schwartz, J. & Jessell, T. (1995) *Essentials of neural science and behavior.* Appleton & Lange. [CDV]
Kay, P. & Berlin, B. (1997) Science ≠ imperialism: There are non-trivial constraints on color naming. *Behavioral and Brain Sciences* 20(2):196–201. [PK, BS]
Kay, P. & Maffi, L. (in press) Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist.* [PK]
Kay, P. & McDaniel, C. K. (1978) The linguistic significance of the meanings of basic color terms. *Language* 54:610–46. [BM, aSEP]
Kevan, P. G. & Backhaus, W. (1998) Color vision: Ecology and evolution in making the best of the photic environment. In: *Color vision - perspectives from different disciplines,* ed. W. Backhaus, R. Kliegl & J. S. Werner. De Gruyter. [WB]
Kim, J. (1984) Concepts of supervenience. *Philosophy and Phenomenological Research* 65:153–76. [aSEP]
Koffka, K. (1935) *Principles of Gestalt psychology.* Harcourt, Brace. [aSEP]
Köhler, W. (1929) *Gestalt psychology.* G. Bell & Sons. [aSEP]
Kranda, K. & King-Smith, P. E. (1979) Detection of coloured stimuli by independent linear systems. *Vision Research* 19:733–45. [KK]
Krantz, D. H., Luce, R. D., Suppes, P. & Tversky, A. (1971) *Foundation of measurement, vol. 1.* Academic Press. [arSEP]
Laakso, A. & Cottrell, G. W. (1998) How can I know what *you* think? Assessing representational similarity in neural systems. In: *Proceedings of the Twentieth Conference of the Cognitive Science Society.* Erlbaum. [HH]
Land, E. H. (1986) Recent advances in retinex theory. *Vision Research* 26:7–21. [BS]
Land, E. H. & McCann, J. J. (1971) Lightness and retinex theory. *American Journal of the Optical Society of America* 61:1–11. [rSEP]
Laws, G., Davies, I. & Andrews, C. (1995) Linguistic structure and non-linguistic cognition: English and Russian blues compared. *Language and Cognitive Processes* 10:59–94. [GVP]
Levine, J. (1983) Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64:354–61. [aSEP]
Liberman, A. L., Harris, K. S., Hoffman, H. S. & Griffith, B. C. (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54:358–68. [aSEP]
Locke, J. (1690/1987) *An essay concerning human understanding.* Clarendon Press. [KK, BM, EM, aSEP]
Lutze, M., Cox, N. J., Smith, V. C. & Pokorny, J. (1990) Genetic studies of variation in Rayleigh and photometric matches in normal trichromats. *Vision Research* 30(1):149–62. [NB]
Lycan, W. G. (1996) *Consciousness and experience.* MIT Press. [NB, AB]
MacLaury, R. E. (1986) Color in Mesoamerica, vol. II: Primary, derived, and desaturated categories. Manuscript in possession of author. [REM]
(1991) Exotic color categories: Linguistic relativity to what extent? *Journal of Linguistic Anthropology* 1:26–51. [REM]
(1997a) *An essay concerning human understanding.* Clarendon Press. [aSEP]
(1997b) *Color and cognition in Mesoamerica: Constructing categories as vantages.* University of Texas Press. [RMF, PK, REM, GVP]
(1997c) Ethnographic evidence of unique hues and elemental colors. *Behavioral and Brain Sciences* 20:202–203. [REM]
MacLennan, B. J. (1995) The investigation of consciousness through phenomenology and neuroscience. In: *Scale in conscious experience: Is the brain too important to be left to specialists to study?,* ed. J. King & K. H. Pribram. Erlbaum. **(http://www.cs.utk.edu/~mclennan/papers/ICTPN-tr.ps.Z).** [BM]
(1996a) The elements of consciousness and their neurodynamical correlates. *Journal of Consciousness Studies* 3:409–24. Reprinted in: *Explaining consciousness: The hard problem,* ed. J. Shear. MIT Press. [BM] **(http://www.cs.utk.edu/~mclennan/papers/ECNC.ps.Z)**
(1996b) Protophenomena and their neurodynamical correlates. University of Tennessee Computer Science Department Technical Report UT-CS-96–331. [BM] **(http://www.cs.utk.edu/~mclennan/papers/PNC.ps)**
(1999) The protophenomenal structure of consciousness with especial application to the experience of color. University of Tennessee Computer Science Department Technical Report UT-CS-99–418. [BM] **(http://www.cs.utk.edu/~mclennan/papers/PSClong.ps.Z)**
MacLeod, D. I. A. & Lennie, P. (1976) Red-green blindness confined to one eye. *Vision Research* 16:691–712. [SEP]
Malcolm, N. L. (1999) Grammars rule OK. *Behavioral and Brain Sciences* 22(4):722–23. [NLM]
(manuscript) De coloribus: Analysis and metaphysics. [NLM]
Marcel, A. J. (1983) Conscious and unconscious perceptions: An approach to the



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

relations between phenomenal experience and perceptual processes. *Cognitive Psychology* 15:238–300. [aSEP]
Marks, L. E. (1978) *The unity of the senses.* Academic Press. [JTT]
Matsuzawa, T. (1985) Colour naming and classification in a chimpanzee (*Pan troglodytes*). *Journal of Human Evolution* 14:283–91. [CLH]
Maudlin, T. (1989) Computation and consciousness. *Journal of Philosophy* 86:407–32. [JS]
Maund, J. B. (1993) Representation, pictures and resemblance. In: *New representationalisms: Essays in the philosophy of perception,* ed. E. Wright. Avebury Press. [EW]
McDowell, J. (1994a) *Mind and world.* Harvard University Press. [EW]
(1994b) The content of perceptual experience. *Philosophical Quarterly* 44:190–205. [NB]
Meadows, J. C. (1974) Disturbed perception of colors associated with localized cerebral lesions. *Brain* 97:615–32. [aSEP]
*Merriam-Webster's Collegiate Dictionary, 10th edition* (1995). Merriam Webster, Inc. [MHB]
Missa, J.-N. (1993) La philosophie face aux donnees neuroscientifiques de la vision des coulers. In: *La couleur,* ed. L. Couloubaritsis & J. Wunenberger. Ousia. [RMF]
Moss, A. E., Davies, I., Corbett, G. & Laws, G. (1990) Mapping Russian color terms using behavioral measures. *Lingua* 82:313–32. [RMF]
Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 83:435–50. [HH, MN- R]
Neitz, J. & Jacobs, G. (1986) Polymorphism of long-wavelength cone in normal human color vision. *Nature* 323:623–25. [NB]
Neitz, J., Neitz, M. & Jacobs, G. (1993) More than three different cone pigments among people with normal color vision. *Vision Research* 33(1):117–22. [NB]
Neitz, M. & Neitz, J. (1995) Numbers and ratios of visual pigment genes for normal red-green color vision. *Science* 267:1013–16. [REM]
Newton, I. (1704/1952) *Opticks (Books 3).* Smith & Walford. [aSEP] Dover. [JvB]
Nida-Rümelin, M. (1996) Pseudonormal vision: An actual case of qualia inversion? *Philosophical Studies* 82:145–57. [KK, MN-R, aSEP, PWR]
(1998) On belief about experience. *Philosophy and Phenomenological Research* 51(1):51–73. [MN-R]
(1999) Psuedonormal vision and color qualia. In: *Tucson III: Towards a theory of consciousness,* ed. S. Hameroff, A. Kaszniak & D. Chalmers. MIT Press. [MN- R]
Palmer, S. E. (1978) Fundamental aspects of cognitive representation. In: *Cognition and categorization,* ed. E. Rosch & B. Lloyd. Erlbaum. [GJO, arSEP]
(1999) *Vision science: Photons to phenomenology.* MIT Press. [arSEP]
Paramei, G. V. & Cavonius, C. R. (1997) Color naming in dizygotic twin protanopes at different luminance levels. In: *AIC Color 97: Proceedings of the 8th Congress of the International Colour Association, vol. 2.* The Color Science Associaion of Japan. [GVP]
Pauen, M. (1999) *Das Rätsel des Bewußtseins. Eine Erklärungsstrategie.* Mentis. [MP]
Piaget, J. (1970) *Genetic epistemology.* Columbia University Press. [EW]
Piantanida, T. P. (1974) A replacement model of X-linked recessive colour vision defects. *Annals of Human Genetics* 37:393–404. [MN-R]
Pielot, R. & Backhaus, W. (submitted) Simulations of coevolution of color vision systems of pollinating insects and spectral reflectance of flowers. II. *Biological Cybernetics.* [WB]
Pielot, R., Breyer, J. & Backhaus, W. (in press) Simulations of coevolution of color vision systems of pollinating insects and spectral reflectance of flowers. I. *Biological Cybernetics.* [WB]
Poincaré, H. (1952) *Science and hypothesis.* Dover. [aSEP]
Pöppel, E. (1995) *Lust und Schmertz. Über den Ursprung der Welt im Gehirn.* Goldmann. [MP]
Posner, M. I. & Raichle, M. E. (1994) *Images of mind.* Scientific American Books. [aSEP]
Proust, M. (1934) *Remembrance of things past.* Random House. [NB]
Putnam, H. (1960) Minds and machines. In: *Dimensions of mind,* ed. S. Hook. York University Press. [aSEP]
(1965) Brains and behavior. In: *Analytic philosophy,* ed. R. J. Butler. Blackwell. [aSEP]
Rainville, P., Duncan, G. H., Price, D. D., Carrier, B. & Bushnell, M. C. (1997) Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science* 277:968–71. [MP]
Ramachandran, V. S. & Blakeslee, S. (1998) *Phantoms in the brain.* Morrow. [JSm]
Rey, G. (1997) *Contemporary philosophy of mind.* Blackwell. [NB]
Riley, C. A. (1995) *Color codes.* University Press of New England. [CLH]
Rommetveit, R. (1974) *On messaage structure: A framework for the study of language and communication.* Wiley. [EW]
(1983) On negative rationalism in scholarly studies of verbal communication and dynamic residuals in the construction of human intersubjectivity. In: *The social contexts of method,* ed. M. Brenner, P. Marsh & M. Brenner. Croom Helm. [EW]
Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In: *Cognitive development and the acquisition of language,* ed. T. E. Moore. [aSEP]
(1975) Cognitive reference points. *Cognitive Psychology* 7(4):532–47. [aSEP]
Russell, B. (1918) *Introduction to mathematical philosophy.* Allen and Unwin. [JSm]
Sahlins, M. (1976) Colors and cultures. *Semiotica* 1:1–22. [RMF]
Sandell, J. H., Gross, C. G. & Bornstein, M. H. (1979) Color categories in macaques. *Journal of Comparative and Physiological Psychology* 93(4):626–35. [aSEP]
Saunders, B. A. C. & van Brakel, J. (1997) Are there non-trivial constraints on colour naming? *Behavioral and Brain Sciences* 20:167–228. [arSEP, GVP]
Schutz, A. (1962) *Collected papers, vol. I: The problem of social reality.* Martinus Nijhoff. [EW]
Searle, J. R. (1980) Minds, brains, and programs. *Behavioral and Brain Sciences* 3(3):417–57. [NB, aSEP, JS]
Sejnowski, T. J. & Rosenberg, C. (1987) Parallel networks that learn to pronounce English text. *Complex Systems* 1:145–68. [GJO]
Sellars, R. W. (1922) *Evolutionary naturalism.* Open Court. [rSEP, EW]
(1932) *The philosophy of physical realism.* Macmillan. [EW]
(1916/1969) *Critical realism.* Russell and Russell. [EW]
Sheinberg, D. L. & Logothetis, N. K. (1997) The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences USA* 94:3408–13. [aSEP]
Shepard, R. N. (1962a) The analysis of proximities: Multidimensional scaling with an unknown distance function: Part I. *Psychometrika* 27(3):125–40. [aSEP]
(1962b) The analysis of proximities: Multidimensional scaling with an unknown distance function: Part II. *Psychometrika* 27(3):219–46. [aSEP]
(1982) Geometric approximations to the structure of musical pitch. *Psychological Review* 91:117–47. [aSEP]
Shepard, R. N. & Chipman, S. (1970) Second-order isomorphism of internal representations: Shapes of states. *Cognitive Psychology* 1:1–17. [aSEP]
Shepard, R. N. & Cooper, L. A. (1975) Representation of colors in normal, blind, and color-blind subjects. Paper presented at the Meeting of the American Psychological Association, Chicago, September 1975. [GRL]
Shoemaker, S. (1975) Functionalism and qualia. *Philosophical Studies* 27:291–315. [NB, RVG]
(1981) Absent qualia are impossible. *Philosophical Review* 90:581–99. [RVG]
(1982) The inverted spectrum. *The Journal of Philosophy* 79(7):357–81. [NB] Reprinted in: *The nature of consciousness,* ed. N. Block, O. Flanagan & G. Güzeldere. MIT Press, 1997. [AB]
(1994) Self-knowledge and inner sense. Lecture III: The phenomenal character of experience. *Philosophy and Phenomenological Research* 54(2):291–314. [NB]
Sloan, L. L. & Wollach, L. (1948) A case of unilateral deuteranopia. *Journal of the Optical Society of America* 38:501–509. [aSEP]
Smythies, J. R. (1959–60) The stroboscopic patterns. *British Journal of Psychology* 50:106–16, 305–25; 51:247–55. [JSm]
(1994a) Requiem for the identity theory. *Inquiry* 37:311–29. [JSm]
(1994b) *The walls of Plato's cave.* Avebury Press. [JSm]
(1999) Consciousness: Some basic issues. A neurophilosophical perspective. *Consciousness and Cognition* (in press). [JSm]
Stalnaker, R. (in press) Comparing qualia across persons. *Philosophical Topics.*[AB]
Stevens, S. S. (1951) Mathematics, measurement, and psychophysics. In: *Handbook of experimental psychology,* ed. S. S. Stevens. Wiley. [aSEP]
Stillings, N. A., Feinstein, M. H., Garfield, J. L., Rissland, E. L., Rosenhaus, D. A., Weisler, S. E. & Baker-Ward, L. (1987) *Cognitive science: An introduction.* MIT Press. [JvB]
Stoerig, P., Barbur, J. L., Sahraie, A. & Weiskrantz, L. (1994) Discrimination of chromatic stimuli in blindsight: Pupilometry and psychophysics. *Investigative Ophthalmology and Visual Science* 35:1813. [PJB]
Stoerig, P. & Cowey, A. (1992) Wavelength sensitivity in blindsight. *Brain* 115:425–44. [EM]
Stwertka, S. A. (1993) The stroboscopic patterns as dissipative structures. *Neuroscience and Biobehavioral Reviews* 17:68–78. [JSm]
Tarski, A. (1954) Contributions to the theory of models: I. and II. *Indigationes Mathematicae* 16:572–88. [arSEP]
Taylor, J. R., Mondry, H. & MacLaury, R. E. (1997) A cognitive ceiling of eleven basic color terms. In: R. E. MacLaury, *Color and cognition in Mesoamerica: Constructing categories as vantages,* (Appendix IV). University of Texas Press. [GVP]
Thompson, E. (1995) *Colour vision: A study in cognitive science and the philosophy of perception.* Routledge. [aSEP]
Thompson, E., Palacios, A. & Varela, F. J. (1992) Ways of coloring: Comparative


https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press

color vision as a case study for cognitive science. *Behavioral and Brain Sciences* 15(1):1–74. [aSEP]

Tye, M. (1995) *Ten problems of consciousness: A representational theory of the phenomenal mind.* MIT Press. [NB, AB, JC]

Uchikawa, K. & Ikeda, M. (1987) Color discrimination and appearance of short-duration, equal-luminance monochromatic lights. *Journal of the Optical Society of America A* 4:1097–1103. [GVP]

Van Gulick, R. (1985) Physicalism and the subjectivity of the mental. *Philosophical Topics* 13:51–70. [RVG]

(1991) Nonreductive materialism and the nature of intertheoretical constraint. In: *Emergence and reduction,* ed. A. Beckermann, H. Flohr & J. Kim. De Gruyter. [RVG]

VESA-Video Electronic Standards Association (1998) *Flat panel display measurement standard,* Version 1 (May 15, 1998), Section A203, Luminance - L(z) for a Diffuse Object. [MHB]

Vollrath, D., Nathans, J. & Davis, R. W. (1988) Tandem array of human visual pigment genes at Xq28. *Science* 240:1669–72. [KK]

Von Glaserfeld, E. (1982) An interpretation of Piaget's constructivism. *Réview Internationale de Philosophie* 142(3):612–35. [EW]

Wallach, H. (1948) Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology* 38:310–24. [rSEP]

Watson, J. D. & Crick, F. H. C. (1953) Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature* 227(4356) 171:737–38. [aSEP]

Weiskrantz, L. (1986) *Blindsight: A case study and implications.* Oxford University Press. [aSEP]

Wendler, D. (1996) Locke's acceptance of innate concepts. *Australasian Journal of Philosophy* 74(3):467–83. [BS]

White, C., Lockhead, G. R. & Evans, N. J. (1977) Multidimensional scaling of subjective colors by color-blind observers. *Perception and Psychophysics* 21:522–26. [GRL]

Wittgenstein, L. (1953) *Philosophical investigations.* Macmillan. [NB, aSEP, JvB] Section 258 in Blackwell edition. [NLM]

Wright, E. L. (1986) Dialectical perception: Lenin and Bogdanov on perception. *Radical Philosophy* 43:9–16. [EW]

(1990) New representationalism. *Journal for the Theory of Social Behaviour* 20(2):65- 92. [EW]

(1996) What it isn't like. *American Philosophical Quarterly* 23(1):23–45.

Zadeh, L. A. (1975) Fuzzy sets and their applications to cognitive and decision processes. Paper presented at the U. S.-Japan Seminar on Fuzzy Sets and Their Applications, University of California, Berkeley. [aSEP]



https://doi.org/10.1017/S0140525X99302211 Published online by Cambridge University Press