13 Consciousness and Isomorphism

Stephen E. Palmer

In this chapter I consider a fascinating problem about consciousness that has intrigued philosophers and scientists since ancient times. Simply put, the question is whether your conscious experiences of color are the same as mine when we both look at the same environmental objects under the same physical conditions. I will call this ontological problem the "color question." I will also consider the important epistemological follow-up question: "... and how could we possibly know?"

The color question is related to an equally old philosophical issue called "the problem of other minds." Here one asks whether organisms or beings other than one's self are conscious or not, ... and how one could know. The color question is not the same as the problem of other minds, in part because the standard position in the color question is to grant that the other being has conscious experiences of color, and to ask only whether those color experiences are the same as one's own under the same conditions. More radical versions of the color question can also be framed—such as the possibility that I have qualitatively different color experiences or even none at all—and I will consider them as well.

The reader may already be wondering why color should be the focus of such a discussion. Why not ask the "pitch question" about sounds or the "saltiness question" about tastes, or whatever might be one's own favorite aspects of sensory experience? Indeed, all of these are perfectly good versions of the same underlying question, which we can call the "qualia question": are my sensory experiences (or qualia) the same as yours or not, ... and how can we know?

Different people have different reasons for focusing on color. My own reason is that we actually know an enormous amount about color perception, and this background of scientific knowledge makes it a good domain in which to ask such questions. I exploited this knowledge in my book (Palmer 1999), in which color vision plays a central role. I use it as the best example of why an interdisciplinary approach to vision is a good idea. Chapter 3 goes through the whole "color story" in detail, all the way from photon wavelengths and retinal cone types to how people in different cultures name colors using basic color terms. It really is a beautiful example. So, when I finally reached the last chapter, which is about visual awareness, I thought an analysis of color might shed some light on the problem of consciousness. And I think it does, in large part because of the huge base of facts that have accumulated over years of scientific research.

Others favor color for historical reasons. In particular, there is a very well known and persuasive argument in the philosophical literature, called the "inverted spectrum argument," that claims to show that we simply cannot know whether or not your color experiences are the same as mine. John Locke advanced this argument in 1690, and it has the following form. There isn't any way you could know whether my experiences of colors are the same as yours or whether they are spectrally inverted. For example, the spatial ordering of my color experiences on viewing the rainbow, going from top to bottom, might literally be inverted relative to yours. If this were the case, you would experience the rainbow with red at the top and violet at the bottom, but I would experience it with violet at the top and red at the bottom. We would both call the top color "red" and the bottom color "violet," of course, because that is what we have all been taught by our parents, teachers, and society at large. Everyone calls blood, ripe tomatoes, and Macintosh apples "red," so we all associate our internal color experiences on viewing these objects—and similarly colored ones—with this verbal label. But might not my internal experiences of color be inverted in just the way

Locke (1690/1987) suggested without its having any effect on how I behave in naming colors? Indeed, Locke claimed that such a spectral inversion of color experience could exist without there being *any* external manifestation, through naming or other observable behavior. It seems there just isn't any way to tell, because I cannot "get inside your head," and "have your experiences," nor can you have mine.

In this chapter I claim that there are ways of rejecting this particular argument without getting inside each other's heads and having each other's experiences. In fact, there is good, solid empirical evidence from behavioral psychology that at least this literal interpretation of Locke's argument is surely false. Once we see why, we can go on to ask whether there is any other transformation of your color experience that I might have without it being detectable in my behavior. There is an interesting generalization revealed by this line of reasoning that leads to an important distinction—which I call the *isomorphism constraint*—between what can and cannot be known about the correspondence of our experiences from behavioral evidence. But before we get to the isomorphism constraint, we need to go back and ground the discussion about color experience in scientific fact to evaluate Locke's argument rigorously.

To begin, we must ask how we could possibly get a scientific handle on a question that asks about the relation between our color experiences. It is pretty obvious that we cannot carry out the thought experiment Locke suggested with real people. What we can do instead is to analyze the inverted spectrum argument from what we know about color science and to see whether any known behavioral data would reveal such an inversion, if it existed.

Color Spaces

One important thing we can measure behaviorally about color experiences is their relative similarities. Everybody with normal color vision agrees, for example, that red is more similar to orange than it is to green. These relative similarities can be obtained for a large sample of triples of colors. It turns out that the results of measuring these three-way similarities can be summarized quite neatly in a geometric model of color experiences known as a *color space*. Each point in a color space corresponds to a color experience, and proximities between points correspond to similarities between colors. This means that nearby points in color space correspond to similar color experiences, and distant points in color space correspond to different color experiences.

Perhaps the simplest and best known color space is Newton's color circle, which is represented in figure 13.1. The saturated colors of the rainbow are arrayed around most of the perime-

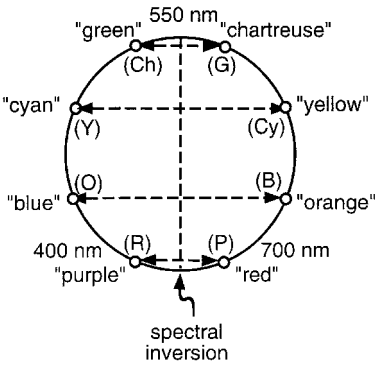


Figure 13.1
Newton's color circle and spectral inversion. Colors are arranged along the perimeter of a color circle, as indicated by the names on the outside of the circle. The dashed diameter indicates the axis of reflection corresponding to literal spectral inversion (rainbow reversal), and the dashed arrows indicate corresponding experiences under this transformation. Letters in parentheses inside the circle indicate the color experiences a spectrally inverted individual would have with the same physical stimuli a normal individual would experience as the colors indicated on the outside of the circle.

ter of the circle. A few wavelengths of light that give rise to these colors are indicated around the outside of the circle, together with English names for a small sample of these colors. This color circle is not the most complete or accurate representation of human color experiences, but it is a good starting point for understanding how behavioral data can constrain the answer to the color question.

One of the interesting things about this geometrical representation of color similarities is that it allows a simple and transparent way to determine whether inverting the spectrum could be detected by behavioral measurements of color similarities. Within the color circle, inverting the spectrum is simply a reflection about the diameter passing through 550 nm, which is approximately the midpoint of the visible spectrum that ranges from 400 to 700 nm. Figure 13.1 illustrates this idea. The color experiences you have are indicated by abbreviations around the outside of the circle, and the ones I have are indicated around the inside. When you experience red (on the outside of this circle), I experience purple (on the inside of this circle); when you experience yellow (outside the circle), I experience cyan (inside the circle); and so forth. So there really is a difference between our color experiences. The dashed arrows in figure 13.1 indicate how our color experiences correspond to each other, a transformation that can be modeled simply by reflection about the indicated spectral inversion axis in color space.

But would these differences be detectible through measures of color similarities? You would say that red is more similar to orange than to green (because the outside point for red is closer to the outside point for orange than it is to the outside point for green). But I would say the same thing, even though, for me, it would correspond to experiencing purple as more similar to blue than to chartreuse (as reflected by proximities of the same points on the inside of the circle). And in fact, all the color similarity judgments you and I would make would be outwardly the same, even though our experiences would be inwardly different. This is just what Locke expected, and it supports his conclusion that spectral inversion would not be detectable.

The reason such differences could not be detected by similarity measures is that the color circle is symmetric with respect to reflection about this axis. We can therefore conclude that so-called spectral inversion of color experiences could not be detected by measurements of color similarity. Furthermore, we can see that this particular transformation is only one of many ways that my color experiences might differ from yours without the difference being detected by measuring color similarities. Any reflection about an axis passing through the center of the color circle would do as well, and so would any rotation about the center. In all these cases, our color experiences would indeed differ, but all our statements about the relative similarities of color samples would be the same.

But there is a great deal more that we can measure behaviorally about color experiences than just their similarities. Among the most important additional factors are relations of color composition, some of which are illustrated in figure 13.2. Most colors look like they are composed of other more primitive colors. Orange, for example, looks like it contains both red and yellow. Purple looks like it contains both red and blue. But there is a particular shade of red that is pure in the sense that it contains no traces of yellow or blue or any other color—it looks "just plain red." And people with so-called normal color vision agree about this fact. Nobody claims, for example, that red actually looks like a mixture of orange and purple, even though it lies between these two colors in color space. Color scientists call these experientially pure colors "unique colors," and there are four of them: unique red, unique yellow, unique green, and unique blue. They are indicated in figure 13.2 by the shaded boxes on the outside of the circle and the color names with boxes around them. All the rest are so-called binary colors.

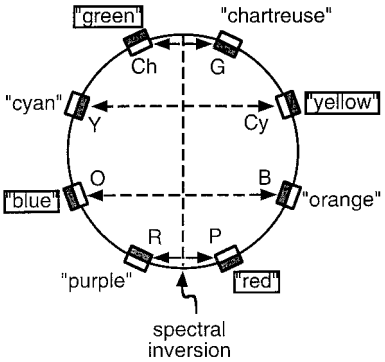
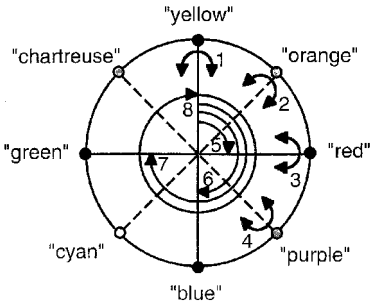


Figure 13.2
Detecting rainbow reversal via unique colors. Shaded rectangles on the outside of the circle represent the four compositionally pure colors (unique red, green, blue, and yellows) for a "normal" trichromat. Shaded rectangles on the inside represent the corresponding pure colors to a rainbow reversed individual, who would perceive unique colors at orange, purple, cyan (blue-green), and chartreuse (yellow-green).



Symmetries of the Color Circle

Reflectional:
1. yellow-blue axis
2. orange-cyan axis
3. red-green axis
4. purple-chartreuse axis

Rotational:
5. 90 degrees
6. 180 degrees
7. 270 degrees
8. 360 degrees (identity)

Figure 13.3
Symmetries of the color circle with respect to color similarities and color composition. This diagram indicates the four central reflections and four central rotations over which the structure of the color circle is transformationally invariant.

The existence of these four unique colors provides another behavioral tool for detecting color transformations. Consider spectral inversion again, this time from the perspective of unique hues. Figure 13.2 shows unique color experiences as gray rectangles. You will designate unique hues at red, blue, green, and yellow (where the gray rectangles are on the outside of the circle), but I will designate them at what we call orange, purple, cyan, and chartreuse (where the gray rectangles are on the inside of the circle). The reason is simply that the experience of mine that is the same as your experience of unique red, results from my looking at color samples that we all call "purple." So for me, "purple" is a unique hue and "red" is not, whereas for you, "red" is a unique hue and "purple" is not. This behavioral difference can thus be used to unmask a rainbow-reversed individual, if such a person existed.

This example shows that unique hues and other relations of color composition further constrain the set of color transformations that can escape detection. We can now rule out literal spectral inversion in the sense of simply reversing the rainbow. Even so, there are still eight color transformations that will pass all behavioral tests of color similarity and composition with respect to the color circle. They are indicated in figure 13.3 as the four central reflections about the unique hue axes (red-green and blue-yellow) and their bisectors (Transformations 1–4 and the four central rotations of 90, 180, 270, and 360 degrees (Transformations 5–8). All have the crucial property that they map unique hues into other unique hues in addition to preserving relative similarity relations among colors.

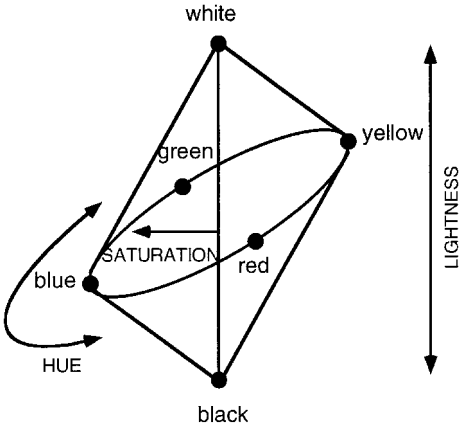


Figure 13.4
Three-dimensional color space. Colors are represented as points in a three-dimensional space according to the dimensions of hue, saturation, and lightness. The positions of the six unique colors (or Hering primaries) within this space are shown by filled circles.

By now, the reader can probably see where this argument is leading. Color transformations that can escape behavioral detection correspond to symmetries in an empirically constrained color space. The important issue for answering Locke's version of the color question, then, boils down to whether there are any symmetries in human color space. If there are, then my color experiences might differ from yours by the corresponding symmetry transformation.

Until now I have been pretending that the color circle, as augmented by the distinguished set of unique hues, is sufficient to represent what is known about human color experience. But there is a great deal more we know about color that is relevant to answering the color question. Most importantly, human color space is actually three-dimensional rather than two-dimensional. The three dimensions are usually called hue, saturation, and lightness, and together they define the lopsided spindle structure diagrammed in figure 13.4. The important fact about the three-dimensional color spindle for purposes of this discussion is that it breaks many of the symmetries in the color circle.

Of most relevance is the fact that highly saturated yellows are quite a bit lighter than highly saturated blues. This asymmetry makes some further color transformations detectable by purely behavioral means. Any transformation in which your experience of yellow is supposed to be the same as my experience of blue (or vice versa) will be detectable because you will say that yellow is lighter than blue, whereas I will say that blue is lighter than yellow (because yellow looks to me like blue does to you, and vice versa). This difference can certainly be detected behaviorally—unless the lightness dimension of my color experience is *also* reversed, so that what looks black to you looks white to me, and what looks white to you looks black to me.

The upshot of such considerations is that if human color space has approximately the structure shown in figure 13.4, there are just three possible color transformations that might escape detection in experiments that assess color similarity and composition relations. They correspond to the three approximate symmetries of human color space shown in figure 13.5. Relative to the so-called normal space in figure 13.4, one transformation (figure 13.5a) reverses just the red-green dimension. The second (figure 13.5b) reverses blue-for-yellow and black-for-white, but not red-for-green. The third (figure 13.5c) is the composition of the other two, which calls for reflecting all three dimensions: red-for-green, blue-for-yellow, and black-for-white.

Although all three are logically possible, the simplest and by far the most plausible is reflecting just the red-green dimension. Indeed, it is so plausible that a good argument can be made that such red-green reversed perceivers actually exist in the population of so-called normal trichromats (Nida-Rümelin 1996). The argument, in a nutshell, goes like this. As figure 13.6 indicates, normal trichromats have three different pigments in their three cone types. Some people are

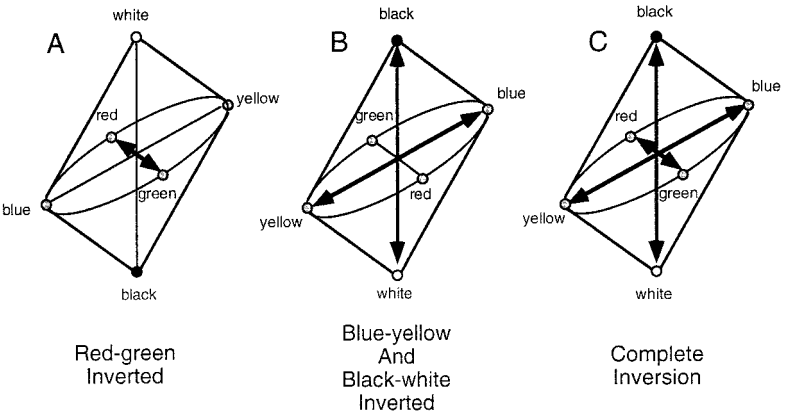


Figure 13.5
Three approximate symmetries of color space. The color space depicted in figure 13.4 has three approximate symmetries: reversal of the red-green dimensions only (*A*), reversal of both the blue-yellow and black-white dimensions (*B*), and reversal of all three dimensions (*C*).

red-green color blind because they have a gene that causes their long-wavelength cones to have the same pigment as their medium-wavelength cones. These people are called protanopes, and their M and L cones are colored gray to indicate that they have the M-pigment in both. Other people have a different form of red-green color blindness because they have a different gene that causes their medium-wavelength cones to have the same pigment as their long-wavelength cones. These people are called deuteranopes, and their M and L cones are colored black to indicate that they both have the L-pigment. In both cases, people with these genetic defects lose the ability to experience both red and green because the visual system codes both by taking the difference between the outputs of these two cone types. But 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. They should exist, and if they do, they are proof that this color transformation is either undetectable or very difficult to detect by purely behavioral means because nobody has ever managed to find one!

There is a great deal more that can be said about the behavioral detectability of color transformations. One key issue for the existence of symmetries in color space is the possible relevance of the basic color terms and basic color categories discovered by Berlin and Kay (1969) in their ground-breaking cross-linguistic studies of color naming. To explain their relevance, I will have to make a brief digression to summarize their findings and theories.

Berlin and Kay made enormous headway in understanding how people describe colors linguistically by restricting their analysis to a basic core of terms. In doing so, they uncovered a very small number of words across all languages that can be used to name all possible colors. They called these words basic color terms (BCTs). BCTs are single, frequently used words that refer exclusively or primarily to colors rather than objects. In English, for example, there are 11

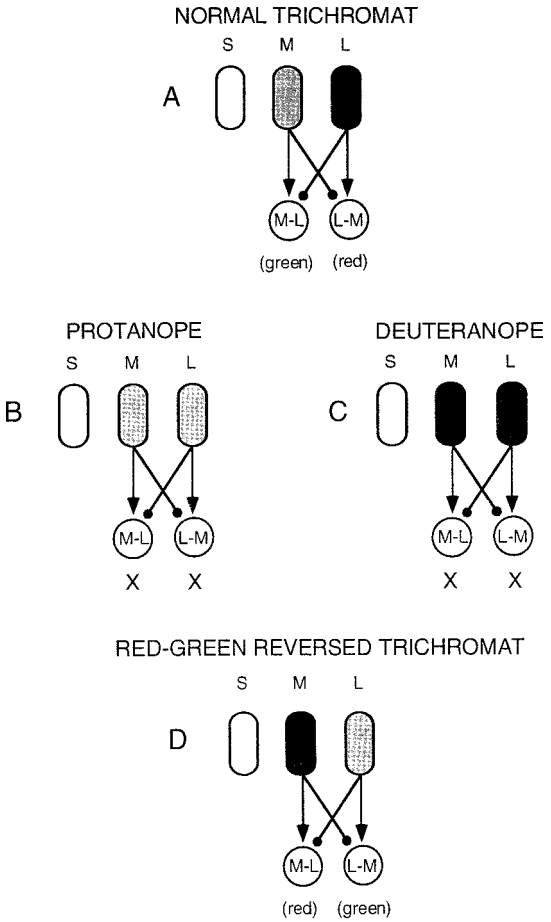


Figure 13.6
A biologically plausible mechanism for red-green inversion. Part *A* shows the "normal" photopigments for short (S), medium (M), and long (L) wavelength sensitive cones. Part *B* shows the result of one form of genetically determined red-green color blindness (both M and L cones have the normal M-cone pigment). Part *C* shows the result of the other form of genetically determined red-green color blindness (both M and L cones have the normal L-cone pigment). Part *D* shows the hypothetical result of both forms of red-green color blindness, which should lead to red-green color reversal.

BCTs: RED, GREEN, BLUE, YELLOW, BLACK, WHITE, GRAY, ORANGE, PURPLE, BROWN, and PINK. (Words like TURQUOISE and SILVER are not included because they refer primarily to substances and only secondarily to colors, and words like CHARTREUSE and CYAN are ruled out because they are not frequent enough.) Still, there are some BCTs that do not appear in English. In Russian, for example, there is a putative BCT (GOLUBOI) for light blue, analogous to PINK in English. In other languages with less fully developed color terms, there are four BCTs that do not appear in English. They can be translated roughly as WARM (yellows, oranges, and reds), COOL (blues and greens), LIGHT-WARM (warm colors plus whites), and DARK-COOL (cool colors plus blacks).

Kay and McDaniel (1978) further analyzed these 16 BCTs into three different types of basic color categories (BCCs), which they called *primary*, *derived*, and *composite*. The most basic are the six primary categories: RED, GREEN, BLUE, YELLOW, BLACK, and WHITE—which they modeled as fuzzy sets with a degree of membership that varies continuously from zero to unity (Zadeh 1965). From these, Kay and McDaniel derived six more categories by the fuzzy-logical AND-ing of two primary color categories:

GRAY is derived from WHITE AND BLACK,

ORANGE is derived from RED AND YELLOW,

PURPLE is derived from RED AND BLUE,

BROWN is derived from BLACK AND YELLOW,

PINK is derived from WHITE AND RED,

GOLUBOI (a Russian word) is derived from WHITE AND BLUE.

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

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

WARM is composed of RED OR YELLOW,

COOL is composed of GREEN OR BLUE,

LIGHT-WARM is composed of WHITE OR WARM which can be defined as WHITE OR RED OR YELLOW,

DARK-COOL is composed of BLACK OR COOL, which can be defined as 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. But it is not clear why there are few or no composite BCTs for RED OR BLUE, GREEN OR YELLOW, WHITE OR COOL, or BLACK OR WARM. These and other mysteries remain to be solved.

The relevance of basic color terms to the inverted spectrum argument is that they may place further behavioral constraints on what color transformations can escape behavioral detection. Many researchers believe that the existence of basic color terms reflects a corresponding set of underlying BCCs into which color experience is naturally partitioned. Because BCTs appear to be linguistically universal, it seems likely that there is something in the underlying structure of human color experience that supports these partitions rather than others that are equally logical. Why are there BCTs for

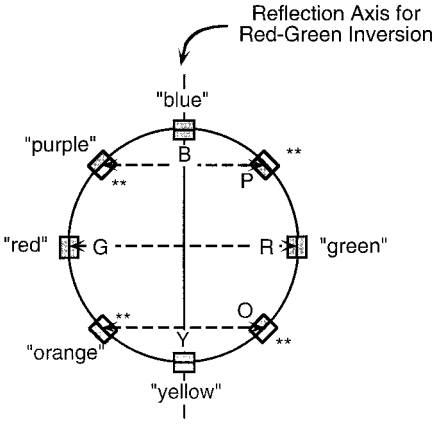


Figure 13.7
Detecting red-green reversal via basic color terms. Shaded rectangles on the outside of the circle represent focal colors of BCTs for a "normal" trichromat, and open rectangles (**) indicate the lack of BCT. The shading of rectangles on the inside of the circle indicates corresponding BCTs for red-green reversed trichromats.

ORANGE and PURPLE, for instance, but not for BLUE-GREEN or YELLOW-GREEN?

Let us now consider the implications of these facts and theories about BCTs for the detection of color transformations via asymmetries in color space. No symmetry problems arise for the six primary BCTs—RED, GREEN, BLUE, YELLOW, BLACK, and WHITE—because they are the same as the six unique colors we have already considered in discussing color composition relations. But if the other ten of Berlin and Kay's basic color terms also arise from singularities in the structure of human color space, then all possible symmetries are broken. Consider, for example, how a red-green invertomat might be detected using the derived BCTs. If I am red-green inverted, I should find it implausible that there are basic color terms for orange and purple, but not for blue-green and yellow-green, as illustrated in figure 13.7. The reason is that my experience of orange is like yours of yellow-green (and vice versa), and my experience of purple is like yours of blue-green (and vice versa). If this asymmetry in basic color categories is rooted in corresponding asymmetries in color experience, I should prefer there to be BCTs for mixtures of what we all call "greens," like cyan and chartreuse (which for me, remember, are experienced as what you would call purple and orange) than for the mixtures of reds (which for me are blue-greens and yellow-greens).

Another asymmetry in BCCs that would unmask a red-green invertomat is the fact that there is a BCT for light-red (PINK) and not for light-green. If I were a red-green invertomat, I should also find this strange, if indeed there is some corresponding asymmetry in color experience. The other two candidate symmetries of color space discussed earlier—complete inversion and blue-yellow plus black-white inversion—are similarly broken by other contrasts where BCTs are not symmetrically arranged in color space.

I am personally not totally convinced that these experiential asymmetries, assuming they exist, would be easily detected in behavior. They seem to be fairly subtle distinctions, and it is conceivable that cultural learning might be strong enough to overpower them. Even if I were a red-green invertomat, the fact that I had been trained all my life with color categories for ORANGE and PURPLE, but not blue-green or yellow-green, might have so firmly changed my thinking about colors that I would not find this way of carving up color experience at all strange, even though I should, at least in principle. Perhaps this is why no red-green invertomats have ever been found, even though they presumably exist.

In any event, the main point of my presentation to this point is that the symmetries of an empirically constrained color space are the key issue in the scientific evaluation of Locke's inverted spectrum argument. I have further

argued that good solid behavioral evidence can be brought to bear on this old philosophical question, and that it rules out all or all-but-three possible transformations, depending on whether one includes BCTs or not. The question I want to turn to now is *why* symmetries of color space are crucial in this argument. This will lead to the second main point of this chapter, which is to identify what I call the "isomorphism constraint" and to discuss its role in the scientific analysis of the color question.

The Isomorphism Constraint

Symmetries have two important structural properties. First, they are what mathematicians call *automorphisms*: they map a given domain onto itself in a one-to-one fashion. This is important for the inverted spectrum argument because one of the ground rules is that both you and I have the same set of color experiences; they are just differently hooked up to the external world. Automorphism is not all that important for the more general color question or other forms of the qualia question, however, because my experience in response to stimulation by wavelengths of light might not be automorphically related to yours. My color space, for example, might be a somewhat shrunken version of yours, such that you would experience colors as more vivid and highly saturated than I do. One might think this would be detectable by the number of jnd's (just noticeable differences) between color pairs, but it wouldn't be if I were simply more sensitive to small differences in my experience than you were, thus compensating for the smaller size of my color space.

More radically, however, we can drop the requirement of automorphism entirely, for my color experiences might be nothing at all like yours. You and I could live in entirely different dimensions of experiential space, so to speak, and it would not matter with respect to what could be inferred about our color experiences from purely behavioral measures. Still more radically, I might have no color experiences at all! I might be a color zombie who processes information about wavelengths of light, yet has no experiences of color whatsoever. (In fact, I know this to be quite untrue of myself, but it might conceivably be true to you!) In any case, if non-automorphic transformations of color experience are allowed, the presence or absence of symmetries within color space becomes irrelevant, and only the other structural property of symmetry matters.

This other property of symmetries is that they are what mathematicians call *isomorphisms*. Isomorphisms are functions that map a source domain onto a target domain in such a way that relational structure among elements in the source domain is preserved by relational structure among corresponding elements in the target domain. In the case of symmetries, the source and target domains are the same (because symmetries are automorphic isomorphisms), but this is not the case for isomorphisms in general. Figure 13.8 illustrates the basic requirements for an isomorphism to hold, using color space as an example. The objects of the source domain (in this case, color experiences) are mapped into those of the target domain (in this case, points in three-dimensional space) so that experiential relations between colors are preserved by corresponding spatial relations between points in color space. This is why spatial models work so well for color experience: They have the same intrinsic structure.

I want to argue that it is isomorphism—"having the same structure"—that is crucial for behavioral equivalence of conscious experiences. This means that as long as two people have the same structure of relations among their color experiences—whatever those experiences might be, in and of themselves—they will always give the same behavioral responses and therefore be behaviorally indistinguishable.

There appears to be a behaviorally defined brick wall, which I will call the subjectivity bar-

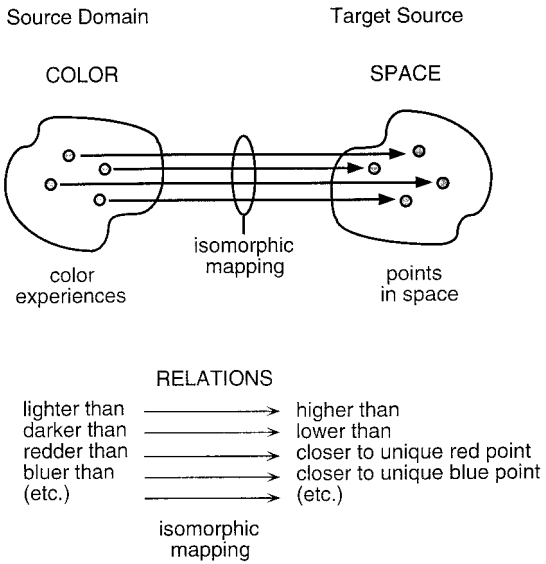


Figure 13.8
The color/space isomorphism. Color experiences are mapped into points in a multidimensional space (see figure 13.4) such that color relations (e.g., *lighter-than*, *redder-than*) are preserved by corresponding spatial relations (e.g., *higher-than*, *closer-to-unique-red-point*).

rier, that limits which aspects of aspects of our experience we can share and which we cannot, no matter how hard we might try. The importance of the isomorphism constraint is that it provides a clear dividing line: the part we can share is the structure in the relations; the part we cannot share is the nature of the experiences themselves. In the case of color experience, this means that we share relational facts such as that red is more like orange than it is like green, that gray is intermediate between black and white, that purple looks like it contains both red and blue, and that there is a shade of red that is compositionally pure. We can share them because they are about the relational structure of experiences. We may implicitly (or even explicitly) believe that we share the experiences too, at least in the sense of supposing that everyone else's are the same as ours, but that does not make it true!

Such arguments suggest to me, as figure 13.9 attempts to depict, that the subjectivity barrier divides shared relations from private experiences and thus coincides exactly with the condition of isomorphism. What I am calling the isomorphism constraint is simply the conjecture that behavior is sufficient to specify experience to the level of isomorphism and not beyond.

The picture that emerges is that the nature of individual color experiences cannot be uniquely fixed by behavioral means, but their structural interrelations can be. In case anyone feels disappointed in this, I hasten to point out that structural relations are absolutely crucial to the fabric of our mental life. Without them, redness would be as much like greenness as it is like orangeness ... or whiteness, or squareness, or middle C, or the taste of pumpkin pie. Without them, perceptual qualities would just be so many equally different experiences, and this certainly is

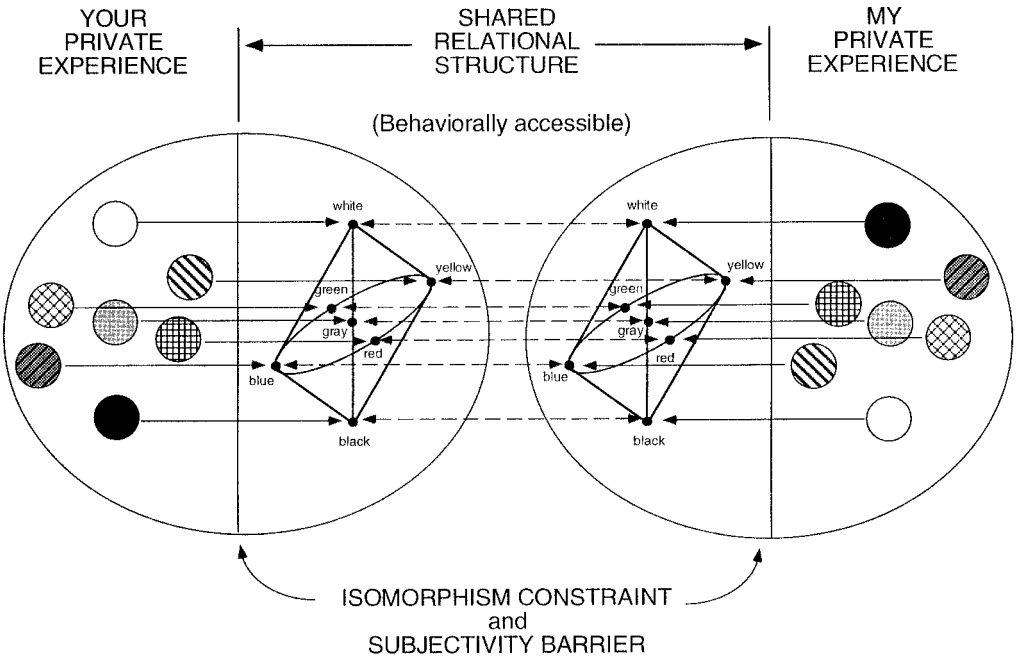


Figure 13.9
The proposed relations among private experiences, shared relational structure, the isomorphism constraint, and the subjectivity barrier in the domain of color perception. Relational structure is publically accessible via observable behavior only to the level of isomorphism. Beyond this level lies the qualitative nature of private experiences.

not so. But, by the same token, structural relations do not reflect everything one would like to know about experiences, for the isomorphism constraint implies that, logically speaking, any set of underlying experiences will do for color, provided they 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 obviously be different.

Behavioral scientists are not alone in working within the constraint of isomorphism, for it also exists in 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 deduction, mathematicians can prove theorems that specify many further relations among mathematical objects within the domain. These theorems are guaranteed to be true if the axioms are true.

But 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 interpreta-

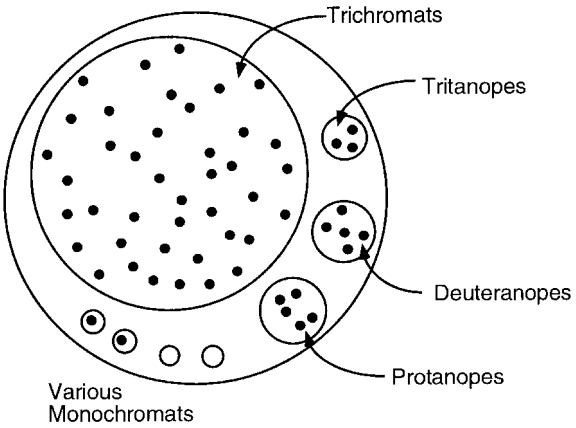


Figure 13.10
Equivalence classes of color perceivers. These Venn diagrams indicate the behaviorally defined equivalence classes of color perceivers who have isomorphic color experiences, but not necessarily equivalent color experiences.

tion 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, and 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 underfined, set of primitives that are free to vary in any way.

The brilliant French mathematician Poincaré (1952) put the situation very clearly. "Mathematicians do not study objects," he said, "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. The isomorphism constraint therefore tells us exactly how far behavioral science can go in specifying experiences.

The Appeal to Biology

If the isomorphism constraint defines the limits of what can be known by behavioral means, figure 13.10 shows where this leaves us with respect to color. Based on psychophysical measures of color science, we can define the standard behavioral equivalence classes of color perception: so-called normal trichromats, three varieties of dichromats (protanopes and deuteranopes, who have slightly different forms of "red-green" color blindness, plus tritanopes, who have "blue-yellow" color blindness), and four types of monochromats. There are some further behavioral classes of so-called color weakness among trichromats that are not represented here, but this classification will do for now.

I have called these behaviorally defined equivalence classes, but with respect to statements about color experiences, it would be more accurate to call them "difference classes." Pairs of

individuals who are in different classes certainly have different color experiences to the same stimulation. Beyond that, we cannot say. There may be many varieties of color experience within the set of normal trichromats, many others within the set of protanopes, and so forth. We just cannot tell on the basis of behavior alone, unless we make some pretty strong further assumptions, such as automorphism, which is difficult to justify at this point in our understanding of color science, given all the differences between different people's visual nervous systems.

This raises the important question of whether there is any way we can go beyond the level of isomorphism by applying biological methods, either alone or in concert with behavioral ones. It is tempting to believe that if consciousness is fundamentally a biological phenomenon, the answer must be, "Of course we can!" I am somewhat less optimistic, but I do not see the situation as completely hopeless, at least in principle, for reasons I will now try to explain.

It seems at first blush that one should be able to study subisomorphic differences in color experiences between two individuals by identifying relevant neurobiological differences and correlating them with differences in color experience, but this will not work. The problem is not in finding biological differences. We will presumably be able to identify the neural differences at whatever level current technology allows. The problem is that, try as we might, we won't be able to identify any subisomorphic differences in experience to correlate with the biological differences. The reason is simply that the subjectivity barrier is still very much in place. Whenever we try to asses how two people's experiences might differ, we can get no further than the isomorphism constraint.

Even so, quite a different line of thought suggests that biology must provide important constraints on the answer to the color question. It seems highly plausible, for example, that two clones, who have identical nervous systems, should have the same color experiences in response to the same stimulation. This is, in effect, a corollary of Kim's (1984) *principle of supervenience*: If the biology is the same, the experiences will be the same. (The converse is not necessarily true, however: If the experiences of two people are the same, the underlying biological processes might be the same, or they might be different.) Most cognitive scientists and neuroscientists ascribe to something like supervenience these days, although it is logically possible that the nature of experience depends on sub-biological facts about quarks, quantum gravity, or some even more esoteric physical entity that has yet to be conceived. I am not going to take such hypotheses seriously until I have to, and will therefore push on with conventional biological approaches, based on the assumption that clones have the same color experiences.

So, assuming the clone assumption to be well founded, is there any way this presumed subisomorphic level of conscious experience can be tapped? The only effective route I can see is one that avoids the subjectivity barrier to some extent by using within-subject designs. The idea is quite simple. Use a biological intervention on an individual and ask for reports about any changes in color experience from before to after the intervention. Suppose, for instance, there were a drug called invertacillin that exchanged the light-sensitive pigments in two classes of retina receptors. Let us also assume that the drug acts reasonably quickly, that it does not mysteriously alter people's long-term memories for object colors, and that it does not disturb the associations between internal experiences and color names. Then it seems plausible to suppose that subjects would indeed notice, and could reliably report, changes in their color experiences after taking the drug. If invertacillin swapped pigments in the medium- and long-wavelength cones, for example, they would presumably report that blood now looks green and grass now looks red. These are extreme examples, and subtler changes in experience would hopefully also

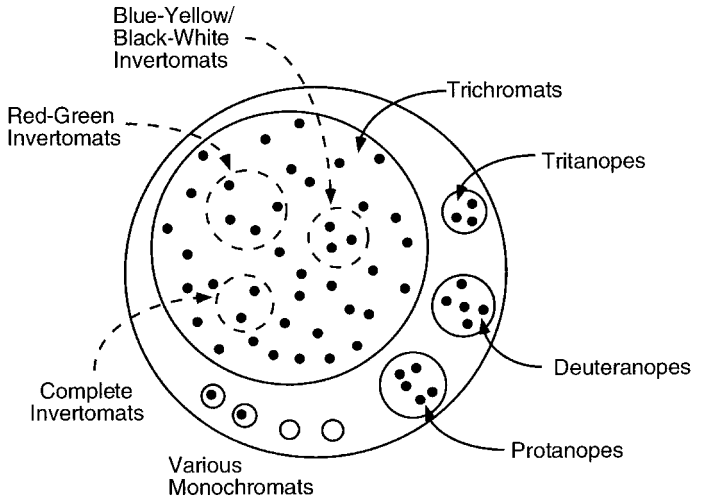


Figure 13.11
Hypothetical subisomorphic classes of color perceivers. Dashed circles indicate the possible existence of three classes of trichromats with color experiences that differ from those of "normal" trichromats at a subisomorphic level, corresponding to the three symmetries of color space indicated in figure 13.5.

be detectable. But the crucial point is that the same subisomorphic color transformations that are difficult or impossible to detect between individuals seem, in principle, quite easy to detect within individuals. Notice that we, as experimenters, have still not penetrated anyone's subjectivity barrier, for we don't actually know how blood or grass appeared to the subject either before or after the change. We only know that it reversed the red-green dimension of color experience, whatever that dimension might be like for that particular observer.

For the sake of argument, let us now suppose that we can figure out what the biological effects of the drug are and that it affects everyone's color experiences in the same way: namely, by reversing the red-green dimension of color space. Armed with this information, we can then divide the set of behaviorally defined trichromats into those who normally have the biological structure associated with the result of the invertacillin intervention (labeled "Red-Green Invertomats" in figure 13.11) and those who do not. Notice that this biologically defined equivalence class does not imply equivalent color experiences for individuals within it. With respect to color experiences, they constitute a difference class, just like behaviorally defined difference classes. People in different difference classes have different color experiences, but people in the same difference class may or may not have the same color experiences. We cannot know whether people in the same class have the same experiences until we exhaust the set of all the relevant biological factors and all their possible interactions, which is a very large set indeed.

But suppose, for the sake of argument, that we could determine the complete catalog of the biological factors that are relevant to color experience in this way. Then we could, in principle, define real equivalence classes of people who presumably have the same color experiences.

Notice that such statements would always be inferences about two people having the same experiences based on certain assumptions, much like our earlier inference that two clones would have the same experiences based on knowledge that their biology is the same. We have plausible scientific reasons to believe that they would, but no way of testing it directly because of the subjectivity barrier. The clones themselves can neither confirm nor deny the conjecture, of course, because the subjectivity barrier exists for them as much as for everyone else.

But, if we were able to carry out this research program—and that may be too big an "if" for anyone but philosophers to swallow—it seems that we would, in principle, be able to infer what colors look like to at least some other people with reasonable certainty. People who are in the same biological equivalence class as yourself would experience the world pretty much as you do, within some reasonable margin of error. And people who are in at least some different equivalence classes might be inferred to have color experiences that differ from yours by identified transformations. If I am a red-green invertomat, for example, and you are a "normal trichromat"—and if the corresponding physiological difference were the only one in our chromatic neurobiology—then our experiences would differ specifically by the red-green inversion transformation caused by the invertacillin drug. You could then know what my color experiences of the world were like simply by taking invertacillin yourself.

But it is important to remember that the possibility that these color-transformed experiences enable you to know what the world looks like to me is necessarily based on inferences. You cannot have my experiences in any direct fashion because of the subjectivity barrier. The inference is based on at least two important assumptions. One is that any differences in experience result from standard biological differences. The other is that all relevant biological variables have been correctly taken into account. If either is false, then the conclusion that you know what it is like to have my color experience by taking invertacillin is also false. Given the dubious nature of at least one of these assumptions, the chances of being able to bring this project off in reality are vanishingly small, even in the long run. Even so, I find the very possibility intriguing.

Acknowledgment

This chapter is based on an article to appear in *The Behavioral and Brain Sciences*. Its writing has been facilitated by grant R01-MH46141 from the National Institute of Mental Health to the author.

References

Berlin, B., and Kay, P. (1969) *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Kay, P., and McDaniel, C. K. (1978) The linguistic significance of the meanings of basic color terms. *Language*, 54, 610–646.

Kim, J. (1984) Concepts of supervenience. *Philosophy and Phenomenological Research*, 65, 153–176.

Locke, J. (1690/1987) *An essay concerning human understanding*. Oxford: Clarendon Press.

Nida-Rümelin, M. (1996) Pseudonormal vision: An actual case of qualia inversion? *Philosophical Studies*, 82, 145–157.

Palmer, S. E. (1999) *Vision science: Photons to phenomenology*. Cambridge, MA: MIT Press.

Poincaré, H. (1952) *Science and hypothesis*. New York: Dover.

Zadeh, L. A. (1965) Fuzzy sets. *Information and Control*, 8, 338–353.