TRUE THEORIES, FALSE COLORS

AUSTEN CLARK

Department of Philosophy
University of Connecticut

Recent versions of objectivism can reply to the argument from metamers. The deeper rift between subjectivists and objectivists lies in the question of how to explain the structure of qualitative similarities among the colors. Subjectivism grounded in this fashion can answer the circularity objection raised by Dedrick. It endorses skepticism about the claim that there is some one property of objects that it is the function of color vision to detect. Color vision may enable us to detect differences in spectral composition without granting us the capacity to detect identities.

1. Objectivism Revivified. An objectivist about color holds, minimally, that colors are mind-independent properties of objects, and that color vision allows us sometimes to detect some of them (Hilbert 1992, 358). Various candidates for these properties have been nominated, and most of them have taken a drubbing. But those initially attracted to the platform, just like those attracted to an "ism" of the ideological variety, can respond to this sorry spectacle in two ways. One is to acknowledge that the best or most likely candidate has been put forth, admit defeat, and find some other "ism" in which to believe. The other is to insist that the best candidate for the cause has yet to be found, urge that none of the previous contenders have accurately embodied the platform, and soldier on.

A similar dialectic can be seen at work in the development of objectivist accounts of color. One aging candidate for what sort of mind-independent property of objects color might be is that of surface spectral reflectance, and as both Hall and Dedrick point out, metamers pose a problem for this identification. Metamers are physically distinct spectra that yield experiences of the same color. Two surfaces that appear to match in color can therefore have very different reflectances.

Various rejoinders might rescue the campaign of this particular candidate, but even if they do not, notice that we can respond to the failure in two ways. One is to agree with Hall and Dedrick that objectivism is a lost cause: color is not a mind-independent property of objects. The other is to change the nominee. Color is not *that* mind-independent property of objects, but perhaps it is some other kind. Invariably this other kind upon which we now pin our hopes must be more complex, to be able to cope with the exigencies discovered in the last campaign. So we get more and more sophisticated forms of objectivism, and more and more elaborate accounts of the sort of property with which colors can be identified.

I think it is worth describing a few steps in this progression. In particular, there are objectivist rejoinders to the argument from metamers presented by both Hall and Dedrick. Suppose we admit that Hall's six different spectra of light each produce in an observer an experience of the same yellow hue, or that Dedrick's two qualitatively identical yellow spots differ in surface reflectance. Must we therefore conclude, as Hall does, that "there is no such property (of physical objects, of light, of anything external) as yellow"?

Those various stimuli all have the power, given the ambient illumination, the observer's state of adaptation, and the angles of view, to produce in the observer an experience of the same yellow. One might try to collect the various wavelength distributions in a disjunctive list. Hall argues that it would be long, open-ended, and unnatural; the disjuncts have "nothing relevant in common". If we posit a dispositional property on this basis alone--a power to cause experiences of yellow--then our attribution is empty, and Dedrick is right to find it circular.

But what if we could give a physical characterization of the relation of metamerism? In fact it can often be explained in terms of retinal processes, specifically in terms of the spectral sensitivity curves and densities in the observer's retina of the three classes of cones: those optimally sensitive to short (*S*), middle (*M*), or long (*L*) wavelengths. Suppose our observer is named Sally. Sally will find two spectra to match whenever those spectra produce equal numbers of absorptions in her *S*, *M*, and *L* cones. They will match because they will have identical effects on Sally's visual receptors, and once they have such identical effects no subsequent developments can differentiate them.

Pairs that are metamers for Sally might not be for Jack, so the relation of metamerism must be relativized to a particular observer--indeed, to a particular observer, in a particular state of retinal adaptation, under particular conditions of illumination, viewing the stimulus from a particular angle. But these are all nice physical variables, possessed by entities that can continue their careers even if no minds accompany them. Eventually we can fill out the terms of a relation of the form "Spectra *A* and *B* match for observer *O* in environment *E*" by predicting quantitatively whether or not *A* and *B* in *E* produce equal numbers of absorptions in *O*'s three cone systems. Indeed the CIE "color matching functions" do something like this for a statistically idealized "standard observer" in precisely defined conditions (Boynton 1979). To predict the matches Sally makes in environment *E*, you would need to know the density of the various cones in Sally's retina, their adaptation state in that environment, the power spectra of the stimuli in question, and the various visual angles and surrounds. But such a color matching function is physicalistically respectable; the details are daunting, but not to a partisan of objectivism. They yield a reply to Hall's question. What property do all six of those spectra share, in virtue of which they all look yellow to Sally? They have in common the property of producing equal numbers of absorptions in Sally's three classes of cones. The disjunction is only as unnatural as the principles of operation of Sally's retina.

The mere existence of metamers implies that color is not an objective property only if the property of yielding equal values on an observer's color matching function is not an objective property. So the debate shifts to the characteristics of the definition of a color matching function (see Hilbert 1987, 110-18). It is defined by reference to effects on a human retina, and so is "anthropocentric". It is a dispositional property. It is relationally defined, since a particular color property would be abstracted from the matching relation. This relation is relativized to particular observers in particular environments. But all of these features have been identified and analyzed in the literature, and sophisticated versions of objectivism in one way or another take account of all of them (Averill 1985, 1992; Broackes 1992; Dretske 1995; Hilbert 1992; Matthen 1988).

The argument from metamers is by no means sufficient to defeat objectivism; it simply prompts a more complex variety. Relationally and anthropocentrically defined dispositions might be defended as "objective", for example. The dispute is not so much over the question of whether there is *any* property that can be identified with yellow, but whether the properties that might fill the bill can be counted as objective or mind-independent.

2. Subjectivism Reconstrued. This dispute is hard to resolve. When some physical phenomenon physically in front of the sense organs looks yellow, almost invariably some property of that phenomenon is cited somewhere in the explanation of why it looks yellow. If we simply add enough qualifications and complexities to the definition of "objective property", and to the account of what sort of objective property color might be, we can ensure that this property, whatever it is, will meet all our criteria. Hire a philosopher to define the criteria of your success and you cannot fail. A more compelling argument for subjectivism would step above piecemeal attacks on the particular candidates, and explain in principle why the objectivist platform fails. It need not even deny the thesis that if you are willing to swallow all those additional qualifications and complexities in the notion of an objective property, then you can hold that *in that sense* color is a property of objects. Its point is rather that there is no need to swallow all of those additional qualifications and complexities. A simpler alternative--a lighter meal--is available. As Hardin puts it

> It is not as if there were no plausible alternative to all these Ptolemaic epicycles. There is, and it is simply this: render unto matter what is matter's. Physical objects seem colored, but they need not be colored. They do have spectral reflectances and the like, and such properties are sufficient to give us a straightforward and detailed account of the stimuli of color perception. To account for the phenomena of color we need not ascribe any other properties to those stimuli, and we find, furthermore, that when we try to do so, the chief result is obscurity in our understanding and caprice in our tactics. So stop the sun and the stars, and start up the earth. The sun's motions, which we so plainly see, are illusory: the movement is on *our* end. (Hardin 1988, 91)

This argument hinges on principles of explanatory priority and simplicity. How are we to explain the phenomena of color? Are they to be explained predominantly by citing mind-independent properties of objects, or by citing peculiarities of the visual system? The objectivist bets that the locus of such explanations will be extra-dermal; the subjectivist that they will predominantly proceed by citing details of sensory physiology.

The latter case is bolstered by identifying some phenomena of color vision that resist explanation in extra-dermal terms. They concern the structure of qualitative similarities and differences among the colors. For example we have the profoundly revealing phenomenon of hue cancellation: that adding yellow can cancel blue. Cancellation reveals itself in the existence of color complements, complementary after-images, and simultaneous color contrast effects. We find four unitary hues, in complementary pairs, whose mixtures yield the binary hues between them. These are aspects of the structure of color quality space, and even if some artfully articulated properties of objects can be identified with colors, the properties thereby identified seem to be of no help at all in explaining this structure. Indeed, even receptor-based explanations are insufficient. Young, Helmholtz, and Hering all agreed that there were three distinct classes of cones, but the Young-Helmholtz theory of color vision had difficulties explaining hue cancellation--one factor in its eventual demise. Explaining that structure requires a model of post-receptoral organization. Opponency derives from principles that lie deep within the sensory physiology of the percipient.

With subjectivism construed in this fashion it is hard to see how the same problems that beset objectivism could, as Dedrick suggests, apply to subjectivism as well. In explanations of the phenomena of color, where shall we find the principles that do the heavy lifting: in the object seen, or the person seeing? It is hard to see how both loci might prove empty. If extra-dermal physical properties cannot explain the structure of color quality space, then one must seek its explanation within the percipient. To claim that the latter explanations are infected by the same problems is tantamount to claiming that we simply cannot explain that structure.

Dedrick levels what he calls the "circularity objection" against both objectivism and reductionist varieties of subjectivism. Both, he says, face a problem that resembles the old functionalist problem of multiple realization. Instances of different types of reflectances might be instances of the same color, and instances of different types of neural states might be instances of the same sensory state. Reduction allegedly requires type identities. If we are to identify types in the reducing science with some disjunction of types in the reduced science, we need to describe what those disjuncts have in common. There seems to be no way to do this without invoking terms from the reducing science. But this renders any such reduction circular.

The version of subjectivism I have sketched is committed to a reductionist thesis: that we can give neurophysiological explanations for data that are recognizably psychological in origin. The receipt of explanations from one discipline for data from another is, in the broad sense, reduction. The subjectivist holds that the physics of objects outside the skin will not explain much of color vision; the explanatory principles that do the work will do that work within the perceiving subject. Now what sorts of explanations will these be? Not much psychology precedes sensation; our story starts, after all, at the outer-most afferents of the nervous system, and any sort of psychological drama takes at least a few synapses to get going. So if you think that humans can explain some of the phenomena of color vision, and you agree that the physics of phenomena outside the skin are not likely to do the job, then you have little option but to agree that the explanations of these sensory capacities will be physiological. And so you are committed a version of reductionism. Welcome to the club.

But the sort of reduction that is contemplated can be completed in a non-circular fashion. That Sally can make the discriminations she does is our explanandum. The claim of reductionism comes down to the claim that some of those capacities can be explained by appeal to her sensory physiology. I do not see where circularity threatens. There are many different varieties of relationship possible between reduced and reducing theories. In particular, the old idol demanding type identities should have been toppled long ago (see Clark 1980). One discipline can explain data from another even though it employs a distinct typology. Some connecting principles are needed, but identities of types are not required.

Perhaps the procedure Dedrick fears is as follows. We might identify neural states *solely* in terms of their role in discriminations. For example, we might identify "neural states of type *Y*" entirely by their role in Sally's discriminations of yellow things, and then try to use "neural states of type *Y*" in explanations of Sally's capacities to discriminate. This procedure would indeed be illegitimate, but it does not characterize how the disciplines actually proceed. Neurophysiology has its own robust and appallingly intricate vocabulary, none of which has been designed to give any particular assistance to the psychologist. So it was a genuine empirical discovery, worth celebrating, that there exist cells in the lateral geniculate nucleus that respond in just the spectrally opponent fashion previously hypothesized by some psychologists (DeValois, Smith, Kitai, and Karoly, 1958). It was an achievement similar to the telescopic confirmation of the existence of an eighth planet, based on the anomalies in the orbit of the seventh. Neither achievement can be charged to linguistic sleight of hand.

Dedrick raises some particular objections to my (1993) account of how one goes about detailing the structure of a quality space, defining places within it, then providing a neurophysiological interpretation for some of its axes. Does this procedure define identity of neural states in terms of places within the quality space? If so, it cannot subsequently proceed to explain the structure of quality space in terms of neural states so defined.

I agree with Dedrick that such a procedure would be circular, but I do not agree that subjectivism is committed to following any such route. In particular, suppose we find places *A* and *B* in the quality space, corresponding to the qualities presented by stimuli that are mutually indiscriminable, yet which can be differentiated, since exactly one of them is discriminable from some third place *C*. Are we committed to the view that there is a distinct type of neural state corresponding to *A*, and a type to *B*, and no type in between?

Well, briefly, no. Let me also disavow any commitment to the existence of "minimum qualitative differences"--what a psychologist would call "absolute thresholds". Talk about "quality space" is a highly condensed (and admittedly confusing) way of talking about the structure of similarities and differences among the qualities presented by stimuli. The key to keeping it straight is that discriminations among stimuli are used to order the qualities that those stimuli present. To find qualities in the *A*, *B*, *C* ordering discussed by Dedrick, the discrimination data used are invariably statistical in character. That is, sometimes stimuli presenting qualities *A* and *B* will be judged by an observer to match, sometimes not. To place qualities *A* and *B* in an order, we must make a statistical characterization of a series of trials. Similarly the difference between qualities *A* and *C* would not be manifest in a single trial, but only over a series; and in some large proportion of trials the stimuli presenting those qualities would be judged to match one another.

It seems reasonable to expect that neural identifications of sensory qualities will also have a statistical character. A quality will correspond to some statistical and distributed characteristic of a population of neurons. Even if there is some absolute minimum to the difference between two individuals, the means of two populations can differ by less. For example, the difference between average spiking frequencies of two populations of neurons can be less than a millisecond, even though each neuron may require a millisecond between spikes. So if qualitative differences are statistical, one is hard-pressed to identify a minimum possible difference.

Perhaps the qualities *A*, *B*, *C* will ultimately be identified with statistical distributions of neural activity--in this case, distributions that overlap. As long as we have a robust physiological vocabulary for talking about the behavior of a single neuron, these statistical characterizations are unproblematic, and we can proceed to describe states of populations of them without falling into any vicious circles.

3. False Colors. Enough of disagreements; now it is time to note the broader areas on which we agree. Hall and Dedrick say some very interesting things about the question of whether color vision has the job of detecting or representing properties of objects. The "representation-alist" or "detectionist" thesis is that there are properties of objects identical to colors, and that color vision detects or represents those properties. Both Hall and Dedrick challenge both conjuncts.

We detect or represent a property if we detect or represent that wherein two instances are the same. Is color vision in that line of business? The argument over metamers itself illustrates the difficulties in being definite about what similarities underlie two patches that look the same color. Color vision is relatively uninformative about the physical similarities that underlie swaths of surface all of the same hue. What is less problematic is the identification of places where the colors differ. Perhaps the critical job is not detection of sameness, but rather the detection of differences. Swaths all of the same hue convey little news. The action is at the edges.

Perceptions of sameness and difference of color show a stunning asymmetry in this respect. Two surfaces that cause experiences of the same color may share no property specifiable independently of the fact that those two surfaces happen to affect our sensory systems in the same way. That is the point of the argument from metamers. But in most naturally illuminated scenes, the places where we see color *borders* typically correspond to places where there is a real physical difference across the border. For example, surface color borders seen under fixed conditions of illumination and retinal adaptation fairly reliably correspond to places where the reflectance *changes*. Our perceptions of sameness do not reliably indicate objective similarities, but our perceptions of surface color differences do reliably indicate physical differences. We should discount the pronouncements of our visual system about sameness, but pay careful heed wherever it finds differences.

Such selective attention may seem inconsistent, but it makes perfect sense whenever we know of biases in judgment. For example, someone who is "bullish" about the stock market will tend almost always to be optimistic, and pronouncements to that effect gradually lose their informative character. But when even the optimists predict a downturn, it is time to pay attention. Our chromatic systems are optimists about similarities. They find them over a vast range of changing reflectance conditions. But when they find a difference--an edge--it is time to pay attention.

Perhaps our wavelength sensitive systems are in the business of detecting differences rather than identities. Consider the primal scene of hunting for a banana in the dense jungle foliage. If we think of the job as property detection, we must identify which reflectance property our color vision reliably detects: that of the banana, or of the foliage. If on the other hand the task is merely to find the banana hiding in the leaves, we do not need to identify properties of the target or of its surround, as long as we can detect some difference between them. The critical task is detecting an edge: a change in the surface properties. Sensed sameness is suspect, but the differences are real.

Satellite telemetry data are often presented in what are called "false color" photographs. Different colors are used to represent arbitrary physical differences in a scene. We might use red to represent reflection of radar waves in the *L* band, and green to represent reflection in the *C* band (see Evans, Stofan, Jones and Godwin 1994). In such a photograph we all admit that the objects on the red side of the border are not really red, and that perhaps the *only* physical similarity that all the items on the red side of the border share is that they affect the detectors in the satellite in the same way. Nevertheless the borders between colors represent real physical differences in the scene. They might show the outlines of ancient craters or where the croplands stop.

Subjectivists claim that all colors are, in this sense, false colors.

Color similarities may not represent any identity of physical properties other than the propensity of disparate classes of objects to affect our receptors in the same way. In that sense, colors are “illusory” (Hardin 1990, 1992). Nevertheless color differences--borders--can be informative about real physical differences in the scene. So false colors can accurately represent the location of real borders.

Illusions can be useful. The sense in which chromatic experiences are illusory in no way precludes their having a useful biological role.

4. Functions of Color Vision. Contemporary objectivists seize on the latter notion, and use it to reply to the banana detection problem. We can identify which of those two reflectance properties--the banana, or the leaves--it is the job of color vision to detect if we can identify the biological function of color vision. The latter in turn requires knowledge of which of those reflectance properties it was advantageous for our ancestors to detect. With this knowledge we could also settle the question of which reflectance property among metamers a sensation of a given color represents (see Dretske 1995, 89-93; Hilbert 1992; Matthen 1988).

I have no quarrel with evolutionary accounts of bio-functions, but I share Hall’s and Dedrick’s skepticism that it can be used in this way to rescue objectivism about colors. The ease with which philosophers can generate plausible hypotheses about the functions of color vision is itself cause for alarm. But a more serious problem lies in the details of mammalian evolutionary history.

Various lines of evidence suggest that our two opponent process systems did not evolve at the same time. The short wavelength photopigment, and the divergence between it and the ancestor of our middle wavelength pigment, are both extraordinarily ancient, dating back over five hundred million years (Goldsmith 1990, 293). Our ancestors in Paleozoic and Mesozoic times who had just two cone pigments were at best dichromats. They could have had only one opponent process system, corresponding to yellow-blue. The divergence between *L* and *M* photopigments, and with it the addition of a red-green opponent system, only occurred within the last sixty five million years, when opportunities opened up thanks to the extinction of the dinosaurs. Trichromacy in mammals is recent and rare. Most species--including the New World monkeys--remained dichromatic. Mammalian color vision still shows the effects of eons of oppression by the reptiles.

Presumably one can tell some story about the selective advantage of dichromacy over monochromatic vision--a story whose climax came over five hundred million years ago. One can tell another story about the advantage of trichromatic vision over dichromatic, which could explain evolutionary events of sixty five million years ago. But there is no logical guarantee that these two stories will be the same. If they differ--if it is even possible that they differ--what then are we to say is “the” function of color vision? If the yellow-blue system was selected for one reason, and the later red-green system for another, what is the function of our sensations of orange? Perhaps we should even think of the two opponent processes as two distinct sensory modalities or modules: one for yellow-blue and a much more recent one for red-green.

As a strategy defending objectivism about color, the dodge into evolutionary history does somewhat resemble the latest Ptolemaic epicycle. Its charms are suspect.

If we think of difference detection rather than property detection we are spared some of these embarrassments. When a species adds a new photopigment, with a spectral sensitivity that differs from those it already had, some previously undetectable differences in the spectral composition of stimuli become detectable. First some equally luminous wavelength distributions can be discriminated, because one looks yellower or bluer than the other. Adding another differentially sensitive pigment adds to the differences in wavelength composition that the organism can detect. Metamers for the dichromat appear as different colors to the trichromat. The trichromat can thereby detect more of the differences in surface reflectance properties. Theoretically there is no upper limit to this process: adding additional pigments would add to the discriminations one could make among wavelength distributions, until finally vision might achieve narrow band-pass filtering, and any difference in the wavelength composition of two stimuli would be detectable. Such a plenitude of differentially sensitive receptors could make vision resemble audition, in which, thanks to narrow band-pass filtering, almost any difference in auditory spectra is detectable. Metamers would more or less disappear. Perhaps then we could be said to detect reflectance spectra.

Unfortunately in visual systems the addition of new differentially sensitive cone systems has a cost: loss of spatial resolution. At some point the costs of lost spatial sensitivity outweigh the benefits of additional spectral discrimination, and one stops with some small number of distinct receptors. They yield the capacity to detect some differences among spectral distributions, and those it fails to discriminate it treats as the same. We get false colors but real differences.

REFERENCES

Averill, E. W. (1985), “Color and the Anthropocentric Problem”, *Journal of Philosophy 82*: 281-304.

--------. (1992), “The Relational Nature of Color”, *Philosophical Review 101*: 551-88.

Boynton, R. M. (1979), *Human Color Vision*. New York: Holt, Rinehart and Winston.

Broackes, J. (1992), “The Autonomy of Color”, in D. Charles and K. Lennon (eds.), *Reduction, Explanation, and Realism*. Oxford: Clarendon Press, pp. 421-65.

Clark, A. (1980), *Psychological Models and Neural Mechanisms*. Oxford: Clarendon Press.

--------. (1993), *Sensory Qualities*. Oxford: Clarendon Press.

DeValois, R. L., Smith, C. J., Kitai, S. T. and Karoly, S. J. (1958), “Responses of Single Cells in Different Layers of the Primate Lateral Geniculate Nucleus to Monochromatic Light”, *Science 127*: 238-9.

Dretske, F. (1995), *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Evans, D. L., Stofan, E. R., Jones, T. D., and Godwin, L. M. (1994), “Earth from Sky”, *Scientific American 271*: 70-75.

Goldsmith, T. H. (1990), “Optimization, Constraint, and History in the Evolution of Eyes”, *Quarterly Review of Biology 65*: 281-322.

Hardin, C. L. (1988), *Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hackett.

--------. (1990), “Color and Illusion”, in W. G. Lycan (ed.), *Mind and Cognition: A Reader*. Oxford: Basil Blackwell, pp. 555-66.

--------. (1992), “The Virtues of Illusion”, *Philosophical Studies 68*: 371-82.

Hilbert, D. R. (1987), *Color and Color Perception: A Study in Anthropocentric Realism*. Menlo Park, CA: Center for the Study of Language and Information.

--------. (1992), “What is Color Vision?”, *Philosophical Studies 68*: 351-70.

Matthen, M. (1988), “Biological Functions and Perceptual Content”, *Journal of Philosophy 85*: 5-27.

Philosophy of Science
PSA Supplemental Issue
63 (3), October 1996, pp. 143-150