Into the neural maze

Donald I. A. MacLeod, UCSD

MacLeod, D. I. (in press). Into the Neural Maze. In: Color Ontology and Color Science. J. Cohen and M. Matthen, eds. Cambridge, Mass, MIT Press.

Recent discussions have pointed to many apparently instructive examples of how neurophysiological discoveries advance our understanding of color. But I believe it is equally instructive to consider the difficulties and obscurities that continue to frustrate us in this project. Here I review some of the presumed successes, and some acknowledged and unacknowledged obscurities, in our current physiological understanding of color, including issues that arise in the investigation of trichromacy; psychological primaries; color discrimination; color blindness; and color constancy. My conclusion is that as we trace the flow of information from the object to the retinal image and thence through the retina and visual pathway, it becomes more and more difficult to discern an isomorphism between color as we perceive it and the neural representation of color in the visual system. Perhaps a closer match to phenomenal experience may be found at higher levels of the system, but the empirical and conceptual basis for that is not yet clear.

Research in visual neurophysiology in the past half-century has produced a wealth of data on neural representations at a variety of levels in the visual system, and there is a general feeling that progress in relating physiology to perceptual experience has been pretty good. After all, Helmholtz in 1867 could still adopt the simplest possible conception of visual processing: that a single nerve fiber runs from each cone photoreceptor to the brain, without interacting with its neighbors, and there produces its associated sensation, like a single pixel in a full-color Cartesian theater. Now we know better. But how might Descartes or Helmholtz have responded to a revelation of the modern discoveries about the visual system? They would surely have been impressed by the new knowledge, but it also seems plausible that they would have been dismayed and bewildered by the seemingly unhelpful complexity and disorder of the physiological substrate of visual experience as we presently understand it. Perhaps they would have considered current ideas about psycho-neural correspondence unrigorous to the point of glibness, finding that instead of clarifying the relationship between brain events and perception, the detailed physiological knowledge we take such pride in only highlights its inscrutability. Whatever Descartes or Helmholtz might have thought, that is the position that I mean to defend here.

The accumulated knowledge about neural representations in retina and brain has never put in doubt the general assumption that our experience of color is directly determined by neural processes. But neither has it yielded well-supported specific linking principles that allow particular aspects of color perception to be ascribed to particular kinds of events. Instead, from the very beginning, new knowledge has brought new difficulties to the concept of psychophysical isomorphism. At the outset, without benefit of any technical knowledge at all, it is natural to suppose that vision makes contact with external objects. This extramission view of perception has found favor among children, among the ancient Greeks, and—notoriously—among present-day students of psychology (Winer, Rader et al. 2003). Given this starting point, the early realization that vision depends on light that travels from its objects to the eye comes as a nasty surprise, forcing the realization that perception is internal to the observer. With the limited physical and neuroscientific knowledge available to

Descartes, the isomorphism between the objective and subjective worlds could still seem fairly straightforward: the perceptual world simply preserves aspects of the spatial and temporal configuration of what's out there, and also specifies information about the values of certain attributes of the external things such as the brightness and color of surfaces. But new problems arose with the recognition that the apparent straightforwardness and immediacy of perception is misleading. Descartes could already trace the beginnings of the causal chain of visual processing in the derivation of the retinal image, the "proximal" stimulus, from the external and distant object. The proximal stimulus is intermediate in the causal chain between the object and the perceptual experience. Yet the perceptual experience is closer to isomorphism with the object than with the proximal stimulus: as perceptual constancies, including color constancy, illustrate, the attributes of perceptual experience are often correlated more closely with properties of their environmental sources than with the proximal signals through which the environment produces its effect (Koffka 1935). This relatively close correspondence is what makes the doctrine of "direct perception" plausible, and has required sense-datum theorists to delve into the depths of experience, without clear success (Thouless 1931; Arend, Reeves et al. 1991), in search of something that has a close correspondence to the proximal stimulus.

Since the retinal image is only two dimensional, depth is absent from the proximal stimulus to vision. In closer inspection, the proximal stimulus reveals other disconcerting features. Instead of being continuous it is fragmented by discrete sampling by the mosaic of retinal photoreceptor cells. So although the proximal stimulus is an intermediate representation of the external world in terms of the causal chain of visual processing, it is not intermediate in form. The stimulus is in fundamental respects less like perception than the original object was.

Newton's demonstration that colored lights generally include all the wavelengths of the spectrum is another nasty surprise, because it disrupts the simplicity of the mapping from object to percept and frustrates realist accounts of color. Some relief comes at the next stage in the causal chain of perception—isomerization of visual pigment—where the infinite spectral degrees of freedom associated with the light stimulus are helpfully reduced to the excitations of only a few types of retinal photoreceptor. But there remains the untidy separation of a single image into three or more different representations that differentially weight the different regions of the spectrum. These must remain segregated, though not completely so, to represent color, but we are left with no answer to the question how the multiple spectrally selective neural representations are integrated to give a unified field of perceived color. This fragmentation continues at postreceptoral stages. "On-center" and "off center" pathways appear at the bipolar cell. The varieties of amacrine and retinal ganglion cells are numbered in the dozens (Kolb, Linberg et al. 1992). Further downstream, cells in primary visual cortex exhibit a practically infinite variety in their functional organization, combining chromatic and spatial selectivity of various sorts, with no indication that any one cell has an independent role in perception, or of what its role would be if it did. After primary visual cortex the progressive complication and fragmentation of the neural representation continues: the signals are dispersed among at least two dozen separate visual areas (Felleman and Van Essen 1991), each forming a woefully distorted map of part or all of the visual field, and none with a very clearly defined role in perception. How might the sporadic firing of the cells that make up this apparently chaotic tangle might form a substrate for perception as we know it subjectively? Surely the honest answer is, we have no idea. We don't know whether or how individual neurons are relevant to perception. An illustrative major advance on this front, following the demonstration that the different visual cortical

areas are functionally different, is the finding (e.g. Salzman et al., 1990; review by Parker and Newsome, 1998) that monkeys' reports about the motion of a stimulus can be altered by concurrent stimulation of neurons in a region concerned with motion. But even this shows only that the neurons in question have some influence on motion perception, and the influence may only be indirect. It is rather like showing that electrical stimulation of the retina can produce a sensation of light: in both cases the neurons in question may be off-stage members of the supporting crew. Instead of solving a pre-existing puzzle the discovery of cortical functional specialization introduces a new challenge, the "binding problem": how do we correctly integrate the motions, colors and other attributes of multiple objects? The explanatory gap between perception and the neural representation as we currently conceive it is wide, then, and new evidence is not making it seem any narrower.

This chapter will survey a few of the "easy" problems (Chalmers 1996) in understanding color perception. The difficulties encountered in the domain of color will highlight the obscurity of psycho-neural isomorphism in general. While homilies on this theme may be abundant enough already, the exercise will at least provide an occasion to review some interesting facts and ideas about color vision.

What makes us trichromatic?

Some colors[1] look the same. In trichromatic vision, we can establish an exact color match using a fixed light and three primaries of adjustable intensity additively combined. The trichromacy of color vision is perhaps the clearest case of successful physiological explanation in perception. According to the standard account, two physically different lights will be subjectively indistinguishable if they are equal to one another in their excitation of the three types of cone photoreceptor that mediate color vision (which we may label as 'L', 'M' and 'S' for long-wave, mid-spectral and short-wave sensitive). The infinite-dimensional variation of spectral reflectance functions is thus replaced by a 3-dimensional variation in neural effect at the photoreceptor stage. The trivariance of photoreceptor signals is a sufficient explanation for trichromacy, and it requires only that the spectral sensitivities of the operative receptors conform exactly to one or other of three possibilities.

Unfortunately, though, this requirement is not generally fulfilled. For one thing, rods provide a fourth visual pigment and hence a fourth degree of freedom for the neural effects of color stimuli. Rods are important for color vision under a wide range of illumination levels. When rods as well as cones are involved, trichromacy is not established at the photoreceptor level. Yet perceptual color matching under these very generally encountered conditions remains trichromatic (Cao, Pokorny et al. 2005); this trichromacy presents a challenge that theory has yet to meet. Further, recent research has identified a fifth potential source of color signals—some varieties of retinal ganglion cell (and perhaps some cone photoreceptors (Dkhissi-Benyahya, Rieux et al. 2006) contain blue-absorbing melanopsin and respond directly to the light that they absorb (as well as conveying signals from the familiar rod and cone photoreceptors) with a spectral sensitivity quite distinct from the rods and cones (Dacey, Liao et al. 2005; Guler, Ecker et al. 2008).

---

[1] Here "color" refers specifically the spectral energy distribution that constitutes the color stimulus, whereas elsewhere I mostly use the word in other meanings. This is the linguistic practice for which Humpty Dumpty has been strongly criticized, but it seems to work perfectly well so long as the shared understanding of context removes serious ambiguity. Many philosophers evidently feel a need for a univocal definition of color—physicalist, subjectivist, dispositional or whatever—but Whittle (2003) encourages us to abandon that dream, and I find his pleading persuasive.

A similarly fundamental complication arises in purely photopic (cone-based) vision for many women, who carry a gene for a 4th cone pigment with an abnormal spectral sensitivity in addition to the three normal ones. Male observers with only the abnormal pigment make "anomalous" trichromatic matches that are often strikingly different from those of normal observers. The mothers of these men generally carry one copy of the gene for the anomaly. They can make trichromatic matches,[2] and their matches are within the normal range, with only a slight deviation toward the anomalous match favored by their offspring. They can, for instance, match a monochromatic yellow in the left half of a circular field to a suitable mixture of spectral red and green in the right half of the field to create the impression of a uniform circular disc. In the carrier women, however, unlike most observers, this match can be upset by adding a uniform long- or short-wavelength veiling light to both sides of the disc (Nagy, MacLeod et al. 1981; MacLeod 1985).

This observation violates Grassmann's third law, which states that matching stimuli can be substituted for one another as constituents of other matches. Since the isomerization of each visual pigment is additive in Grassmann's sense, a match achieved through equality of pigment isomerization rates would satisfy Grassman's third law. It follows that the carrier women's non-additive trichromatic matches are not matches for the pigments. On the standard genetic model (Nathans, Merbs et al. 1992) the carriers have at least four visual pigments: one of the X chromosomes specifies the normal L and M pigments, while the X chromosome from the other parent specifies, instead of L or M, its spectrally deviant "anomalous" counterpart, say L*. The trichromatic matches of these women migrate between the normal match and an "anomalous" one, depending on whether the added veiling light, red or blue, favors use of the anomalous pigment (L*) or the corresponding normal one (L). The ability to involve the different pigments selectively using adaptation to added red or blue light implies that the four pigments are housed in different receptors. This too is expected: women and other female eutherians are a mosaic of maternal and paternally derived X-linked characteristics, following random inactivation of one or the other X chromosomes on a cell by cell basis during development (Mollon 1989).

Since trichromacy in all these cases does not reflect trivariance of the neural response at the photoreceptor level, it must reflect a postreceptoral constraint—a postreceptoral neural trivariance of some kind. But exactly what does this mean? This question is seldom posed, and on close investigation we will discover that there is no well supported answer.

Conceivably, at some postreceptoral stage of processing, the neural signals might fall into only three chromatic classes, having three distinct spectral sensitivities, with practically complete uniformity in spectral sensitivity within each class (Brindley 1970). But this is not consistent with what is known about the spectral sensitivities of single cells in visual cortex. These show considerable individual variation, with almost no agreement among investigators as to how they should be classified; there is no evidence for a discrete set of cell types each having a precise and invariant pattern of connections to the receptors (Lennie, Krauskopf et

---

[2] Part of the fascination of research with these subjects is the intriguing possibility that they might be tetrachromatic and have an added dimension of color experience. Jordan and Mollon (1993) consider this an open question, and some of their experiments do suggest an ability to distinguish between Rayleigh-matched color pairs. The most economical explanation of that finding, however, is that the two fields appear to differ in texture, as would be expected where a trichromatic system has nonuniform spatial properties owing to the interleaving of cones specified by maternally and paternally derived genes. Informal reports of our subjects are consistent with this interpretation. Jameson and Highnote (2001) found that their sample of such women divided the spectrum into more bands than normal; this is consistent with tetrachromacy but is not strong evidence for it.

al. 1990; De Valois, Cottaris et al. 2000). Where in this chaotic diversity of chromatic responsivess can we find a basis for trichromacy? One could suppose that the untidy complex of color signals observed in primary visual cortex is superseded by, or perhaps conceals, a cleanly trichromatic organization: the brain might secrete trichromatic colors in the form of three different neurotransmitters (erythrogen, chlorogen, cyanogen?); or there might turn out to be just three chromatic master neurons at each spatial location, that combine in an orderly way the chaotically diverse chromatic signals that they receive as their inputs. But a more plausible alternative to such extravagant *ad hoc* speculations is that central trichomacy occurs because postreceptoral neural connections treat alike two classes of photoreceptor (for instance the paternal and maternal derived photoreceptors in the carriers of genes for anomalous trichromacy). An *equivalence* of photoreceptor signals in this sense would occur if the 'L' and 'L*' cones are equivalent in their pattern of neural connections.

A parallel but simpler case is provided by the subjective equivalence of right-eye and left-eye stimulation, which can be made subjectively indistinguishable (Barbeito, Levi et al. 1985). Newton guessed that optic chiasm unites left and right fibers that serve a common direction in visual space; this could provide a physiological basis for binocular single vision, and also for the subjective equivalence of vision with the left and with the right eye. But electrophysiology has created a serious difficulty for this view. Although binocularly activated single cells exist in the brain, left and right eye inputs are relayed to separate alternating thin slabs of tissue in thalamus and primary visual cortex. If perception made use of the information available in primary visual cortex, left and right eye stimuli should always look very different. Yet these stimuli, so distinct in their cortical representations, are subjectively indistinguishable. One way to resolve this paradox is to suppose that activity in primary visual cortex has no direct role in perception (Crick and Koch 1995), and that at some later stage of cortical processing binocular convergence becomes complete. This proposal may not carry conviction, as there is no experimental or theoretical reason to think that the binocular convergence is ever complete enough to account for the subjective identity of left and right eye input. And although the proposal requires us to assume that a large proportion of cortical cells—ones with a marked left or right eye preference—have no direct influence on perception, it does not come with a principled criterion for deciding which cells do have direct influence. Still, it is the best story we have. Perhaps, then, the postreceptoral trivariance that forms the basis for trichromacy when more than three visual pigments are involved could analogously originate from a convergence between the incoming signals that practically obliterates the distinction between 'L' and 'L*' inputs?

It turns out, however, that equivalence of the 'L' and 'L*' photoreceptor signals will not guarantee trichromacy. To see why, denote the cone excitations (the quantum catches in the respective visual pigments) by L, M, S and L*. On the equivalence assumption, postreceptoral responses must be uniform: it will not be possible for some postreceptoral responses to depend mainly on L, and others mainly on L*. They can, however, depend on both the sum, L + L*, and the product, LL*, of the quantum catches L and L*. For a match to be valid for both a "sum" neuron and a "product" neuron, both L and L* have to be equal (or swapped) for the two stimuli compared, just as if those quantum catches were independently represented by separate postreceptoral neurons. This makes the system dichromatic. Throw in the 'M' and 'S' signals, and it becomes tetrachromatic. The equivalence hypothesis therefore does not guarantee trichromacy.

Besides, there is good evidence against the equivalence hypothesis as a general principle: it fails in certain doubly dichromatic women who have genes for two L pigments from one parent, and for two M pigments from the other. The color vision of these

"compound heterozygotes" is fully normal despite the demonstrated red/green dichromacy of both parents and/or their male descendants (Carroll 2006). Their trichromacy is surprising because it requires the visual system to segregate the photoreceptor signals that originate from the maternally and paternally derived visual pigments. Suppose we classify the photoreceptors as 'L' or 'M' (in quotes) based on the genetic locus that encodes their pigment (not on the pigment that happens to be encoded there). Each class in the compound heterozygote then has the same proportion of cells with L and M pigment, and if central connections were determined by class membership as we have defined it, dichromacy would result. To explain the compound heterozygote's trichromacy, the fate of her photoreceptor signals must be determined instead by the pigment that they come from. This could happen in two ways: either the photoreceptor cells are labeled, for the purpose of forming connections, by the pigment they contain, or else the patterns of connections are contingent on activity—pigment-dependent activity—during development.

But as we have seen, pigment-based segregation does not occur with carriers of anomalous trichromacy, or they would be tetrachromats. So perhaps the number of labels available to classify photoreceptors for the purpose of forming central connections is genetically limited to three? If so, the basis of post-receptoral trichromacy resides in some kind of trivariance of the molecular biological machinery that allows photoreceptor signals to be segregated into three (and only three) classes in the formation of their central connections.

This idea is close to a restatement of the equivalence hypothesis, but by invoking more specific constraints on the processing of the signals, the idea can be put in a form that implies trichromacy without requiring that central color signals fall into just 3 well-defined classes. A tetrachromatic system will be reduced to a trichromatic one if signals from multiple photoreceptor types are combined *additively* in a fixed manner into a smaller number of postreceptoral signals—a scenario that I will dub "uniform additive combination". Assume that the *i*th member of a representative array of postreceptoral neurons receives an input

$$E_i = f_i (f_L(L) + f_{L^*}(L^*), f_M(M) + f_{M^*}(M^*), f_S(S+S^*)) \quad \text{Equation (1)}$$

Here L and M are the visual pigments from one X chromosome; L* and M* are the pigments from the second X chromosome and apply only to females. Two forms of the S pigment, S and S* will also be present (even in males) but their excitations are shown as additively combined, since both forms are likely expressed in each 'S' photoreceptor. The nonlinear functions f that relate neural signals to one another, or relate photoreceptor signals to pigment quantum catches need not be specified here, but f_L and f_{L^*} are the same for all i (though they can be different from one another, thereby relaxing the equivalence condition), as are f_M and f_{M^*}. Only f_i may be different for the different postreceptoral neurons, but this alone allows the different central neurons to have an infinite variety of chromatic spectral sensitivities, in broad agreement with observation. Despite this diversity of spectral sensitivities, the system does satisfy neural trivariance since the three quantities $f_L(L) + f_{L^*}(L^*)$, $f_M(M) + f_{M^*}(M^*)$ and S determine the inputs to all the neurons at the postreceptoral stage.

Uniform additive combination as embodied in Equation (1) guarantees trichromacy for a system with more than 3 pigments. If the central neurons directly relevant for color perception average across a sufficient number of cones to obtain a fairly representative sample of both maternal and paternal cone signals, this might make trichromatic matches good enough to be acceptable in practice.

These considerations apply not only to carriers of genes for anomalous color vision, but also to many normal women. The normal visual pigments show considerable variation from one person to another (Webster and MacLeod, 1986, Neitz and Neitz, 2000), so even genetically normal women have both their father's and their mother's versions of the normal L and M pigments. The different versions of the L pigment are often different enough in peak absorption to create quite noticeable differences between Rayleigh matches determined by the maternal vs. the paternal L pigments. For most women, therefore, trichromacy is not established at the photoreceptor level: if perception made full use of the information from the photoreceptors, many women would require four or perhaps five primaries to make color matches, depending on whether the maternal/paternal differences between their L pigments, their M pigments, or both are enough to be visually significant. Thus the trichromacy of many women is established postreceptorally, perhaps in the way sketched above. For explaining trichromacy in males when both rods and cones are active, similar considerations apply: trichromacy will hold if rod participation can be modeled by the addition of independent rod terms to each of the three arguments in Equation (1). Finally, the melanopsin ganglion cells could be accommodated in an extension of this theoretical scheme; those neurons, however, are thought to have no influence on color perception at all, although there is no obvious reason for their exclusion from perception given their chromatic responsiveness and cortical connectivity (Dacey, Liao et al. 2005).

The general conclusion here is that we have made almost no progress toward establishing a neurophysiological explanation of trichromacy, or even toward clarifying the logical requirements for such an explanation. Even if the present proposal of uniform additive combination were accepted in principle, its empirical status is quite uncertain.

But trichromacy has a phenomenal as well as a behavioral connotation: a three-dimensional space is generally supposed to be enough to characterize the diversity in appearance of colors, at least for uniform fields viewed in a dark surround. We consider next the neural basis for the qualitative characteristics of perceived color under such reduced conditions.

The psychological primaries

Despite the qualifications just noted, the notion that the photoreceptors provide three unipolar signals for representing color is a useful idealization. But at postreceptoral stages the representation of color undergoes a radical transformation, in which two distinct classes of neurons carry bipolar color-opponent signals, sometimes loosely referred to as a blue/yellow signal and a red/green signal, while a third class is driven by achromatic contrast. At the level of brain processing where input from the optic nerve is received, the three classes are distinct in size and connectivity, and are named respectively konio(-cellular), parvo(-cellular) and magno(-cellular) types (Hendry and Reid 2000).

A recoding of this kind was seen as theoretically attractive as early as the 1890s on the grounds that it would account nicely for the phenomenally unmixed natures of the four psychological primaries, red, green, yellow and blue (also known as unitary, or unique, colors), and the fact that other colors appear to be compounds in the sense that orange, for example, is both reddish and yellowish. The electrophysiological confirmation of that conjecture (Derrington, Krauskopf et al. 1984; De Valois, Cottaris et al. 2000) has been hailed as an example of successful physiological explanation in perception, perhaps second only to the three-receptor account of trichromacy. But here again, the picture becomes

much less clear on close scrutiny. Notably, the "red/green" cells do not correspond well with the psychological primaries (see for instance Abramov and Gordon (Abramov and Gordon 1994)). Whereas ideally they would be unresponsive to colors that are neither reddish nor greenish, they actually respond similarly to green and to blue. They behave in that way because they are driven by a difference between the 'M' and the 'L' cones, and the ratio of the 'M' to the 'L' cone sensitivities is in fact even higher for blue than for green. This has been known or strongly suspected since the time of Helmholtz's pupil, König, and the relevant cone sensitivities are now known with high precision (Stockman, MacLeod et al. 1993). Embarrassingly enough, the stimulus that most strongly excites the midspectral ("green") cones, relative to the long-wave cones, and hence maximally polarizes the "red/green" opponent neurons in the "green" direction, is a distinctly reddish spectral violet, with a wavelength near 455 nm.

Thus a pure blue, devoid of redness and greenness, stimulates both the parvocellular and koniocellular neurons. The implied code for redness is not a signal confined to the parvocellular cells, but a specific combination of parvocellular and koniocellular activation. In view of this it is difficult to maintain that the color-opponent neural code has any functional relation at all to the psychological primaries. This leaves us without a known or plausible account of the psychological primaries. The opponent neural recoding may be useful for reasons unrelated to the psychological primaries. It could be an instance of a frequently encountered principle of functional organization in the visual system, where neural responses are "sharpened", or made more selective, by offsetting excitation with inhibition from nearby inputs. Center-surround antagonism, for instance, is nearly ubiquitous in the spatial receptive fields of sensory neurons, where it enables improvements in spatial resolution. In the color case, it is not the spatial but the spectral sensitivity profile that is sharpened by the arrangement. Perhaps more important, the opponent code improves efficiency by reducing the correlation among the different neural measures of color. And because the opponent signals are minimal for some broad-band gray or near-gray color, the encoding scheme facilitates high sensitivity to small deviations from neutrality. This helps make the relatively desaturated natural colors that are most often encountered in nature highly discriminable, at a small price in reduced discrimination within the seldom frequented extremes of color space (von der Twer and MacLeod 2001).

Some evidence (e.g.He and MacLeod 2001; Shady, MacLeod et al. 2004; Vul and MacLeod 2006) suggests that the earliest cortical stages, where the color-opponent konio and parvo cells deliver their signals, make no direct contribution to color perception. This leaves open the possibility that somewhere along the neural journey to the unknown seat of consciousness, a color opponent code could be created that does correspond to the psychological primaries and can account for the phenomenally unitary nature of red, green, yellow and blue. But alas, there is as yet no sign of such further recoding. On the contrary, the cortical representation of color is so untidy that it has not yet yielded to any simple description (Lennie, Krauskopf et al. 1990; De Valois, Cottaris et al. 2000).

When is a neural substrate simple?

Many color scientists, acknowledging that the color opponent signals observed in the pathway to cortex have no relation to the psychological primaries, do nevertheless take it for granted that a color opponent neural representation capable of accounting for the phenomenally simple or unitary quality of the psychological primaries must exist somewhere

in the brain—in a region that is *directly* reflected in phenomenal experience, instead of merely conveying signals from the eye. This tenet was long maintained in the absence of neurophysiological evidence, and continues to be maintained even though current neurophysiological evidence does not support it. Its *a priori* plausibility derives from convictions made explicit as psychophysical axioms by G. E. Müller (in the translation of Boring, 1942):

"1 The basis of every state of consciousness is a physiological process, to whose occurrence the presence of the conscious state is joined.

2. To an equality, similarity or difference in the sensations...there corresponds an equality, similarity or difference in the psychophysical process, and conversely. Moreover, to a greater or lesser similarity of sensations, there also corresponds respectively a greater or lesser similarity of the psychophysical processes, and conversely.

3. If the changes through which a sensation passes have the same direction, or if the differences which exist between series of given sensations are of like direction, then the changes through which the psychophysical process passes, or the differences of the given psychophysical processes, have like direction.

Moreover, if a sensation is variable in *n* directions, then the psychophysical process lying at the basis of it must also be variable in *n* directions, and conversely."

It is the final point, with its reference to multiple directions, that provides the foundation for a physiological account of the psychological primaries. But on close inspection the meaning and application of the axiom become disturbingly unclear. Some of the difficulties are the following.

First, if, as is generally assumed in the modern story of psycho-neural isomorphism, the different aspects of color are associated with different neural systems each generating a univariant signal, a problem arises in accounting for the subjective integration of the phenomenal dimensions of color. The judgment whether two colors are the same can be made rapidly and with high precision. Deciding whether an orange is redder than a purple is much more problematic. So how can the easy recognition of (multidimensional) color identity depend on the difficult comparison of the individual chromatic dimensions? To answer that those signals are introspectively inaccessible, and are integrated into a unitary percept by some unspecified subsequent processing, would be to give up the desired direct correspondence between perception and the unidimensional physiological signals.

Second, to account for the unmixed simplicity of pure yellow, we want to be able to say that when the addition of a reddish or greenish tint creates a reddish-yellow or greenish-yellow compound color, this happens because some signal for redness or greenness is introduced into the neural representation—a signal that is absent for yellow. The axiom does not allow this because it does not stipulate any zero point for the neural ("psychophysical") processes. Neurons do have a straightforward and functionally meaningful null signal: zero firing rate. This makes it natural to identify Müller's psychophysical process with the firing rates of *n* sets of neurons. But experimentally, a red/green opponent neuron which fires vigorously for red also fires for yellow, albeit at a more leisurely rate—perhaps close to its "spontaneous" rate, the rate observed without any stimulus—and is inhibited from firing for blue, blue-green or greenish-yellow. If such neurons are responsible for phenomenal redness, there should be no uniquely pure and simple yellow: all yellows should be reddish. A related paradox: when neurons fire at their spontaneous rates, presumably nothing is perceived (since that is the neural state associated with an absence of physical stimulation). This is a

simple perceptual outcome, but the corresponding neural state is complex in the sense that we have a full complement of nonzero signals.

Perhaps we might take a different tack by supposing that firing at the spontaneous rate is actually the simplest state of the neural system. But how might one proceed to justify that claim? Perhaps phenomenal simplicity corresponds to a balancing of signals? What, in neural terms, defines balanced signals? Must there be a neural difference signal somewhere in the head, which is zero in the balanced condition? In that case, don't the balanced signals themselves lose their claim to independent and direct phenomenal relevance?

Nevertheless Müller may have been wise to avoid formulating his axioms in terms of explicitly identified signals that range from zero. Although neural firing rates have a straightforward and functionally meaningful zero point, that is not true for many other physiological variables, notably membrane potentials, which are no less plausible than firing rates as contenders for the status of variables of the psychophysical process. This raises the question of the dimensionality *n* of the neural representation. Is there one dimension per neuron? If so is that the neuron's firing rate? If it is, what is the justification for neglecting innumerable other variable parameters of a neuron's state? Do we have to group neurons in classes, and if so should their responses just be averaged if the members of the class are not truly homogeneous in their responses? How could we justify restricting the substrate of color to a circumscribed system of *n* classes of neuron, given that no subset of neurons operates in isolation from the rest of the brain? And perhaps most intriguing: is the direction of the phenomenal change from gray to green truly the same as the direction from red to gray; and if (as I would contend) it is not, can we reconcile this with the observation that the opponent cells provide two bipolar neural dimensions or signals, rather than four monopolar ones?

A third, and related, problem in the search for psycho-neural isomorphism in color vision is that the multiplicity of neurons and the diversity of their chromatic responses gives the neural representation a great excess of dimensions beyond the three under discussion. It is important to recognize that the diversity of receptor spectral sensitivities discussed above is only one reason for the postreceptoral diversity. Even if the photoreceptor spectral sensitivities fell into precisely three classes, the responses of postreceptoral neurons conforming to Equation (1) could in principle take arbitrary values at any point in the three-dimensional space of receptor excitations, because suitable choice of f_i, determined by the intervening connections, can shape a neuron's response to any given color independently of its response to similar colors (except perhaps for a continuity constraint), and this can be done independently for different neurons. The potential diversity of response distributions in cone excitation space is therefore limitless. The responses actually observed in neurophysiological experiments are less chaotic than they might be, but they do, as noted, clearly resist description in terms of a small number of classes (Lennie, Krauskopf et al. 1990; De Valois, Cottaris et al. 2000). Conceivably, some of this neural diversity may be phenomenally relevant. There could be neurons to indicate whether a color is brown, or belongs to some other particular category, or neurons that represent by their firing how warm or cool a color is felt to be. Still, it would be a daunting challenge to find a phenomenal counterpart for the signal from each of a presumably large class of color sensitive neurons in, say, primary visual cortex or later cortical areas specialized for color. But that is what the standard conceptual framework embodied in Müller's axioms leads us to expect.

A final serious difficulty for simple psychoneural isomorphism is introduced by the mixing of neural signals for color and spatial pattern. Rather than being responsive to a

single point in space, visual neurons have a great variety of receptive field profiles. How does the visual system encode both color and form? One possibility is to represent color independently of form, another is to reduplicate the entire apparatus of spatial vision as often as needed to specify color as well as lightness. Neither of these idealizations corresponds to neural reality. There is some functional specialization for color and for form (Livingstone and Hubel 1988) (Zeki 1990) yet many neurons are jointly selective for both, as shown by both subjective phenomena (McCollough 1965) and electrophysiology (e.g. Johnson et al, 2001, 2004). It is not clear how in this situation we can identify a limited set of dimensions for a psychophysical process that represents color in the way envisaged by Müller. If there are clean principles underlying the multiplexing of chromatic and spatial information, they remain unknown, although a beginning is being made in investigating this (Engel, 2005).

These difficulties are avoided an entirely different alternative view of the functional role of visual neurons in color perception, a view that provides a more natural account of the diversity of their chromatic responses. We can give up on *isomorphism*, and require only *consistency* of the responses to support perception. This is tantamount to acknowledging that present neurophysiological data do not yet account for phenomenal trivariance, or for the psychological primaries. But it is more than an admission of failure because it opens the door to other theoretical possibilities, discussed in later sections of this chapter.

Discrimination and similarity

Recall the second of Müller's psychophysical axioms. "To an equality, similarity or difference in the sensations...there corresponds an equality, similarity or difference in the psychophysical process, and conversely." This remains the standard story, and perhaps the best one we have, for the neural basis for similarity and discriminability of colors. But it is problematic almost to the point of incoherence.

- The notion of equality between two instances of the psychophysical process is not well defined. A precise identity at the molecular level is a practical impossibility; and if that is not required, what is?
- As previously noted, we have no principled way of delimiting Müller's "psychophysical process", the neural substrate *directly* relevant to color perception within the brain. It would be pleasing if the identity principle could be applied piecewise to individual modalities or regions, but there is no reason to think that this is justified, nor is it clear how any suggested partitioning could be acceptable given that every candidate neural subsystem is interpreted by processes from neighboring systems.
- Even if we were told which characteristics or "dimensions" of the psychophysical process do need to be considered, we would still have no principled way of determining the similarity of two such processes, because this would require a particular weighting of the differences along each of the given dimensions. A suggestion like "just add the absolute differences in the firing rates of the relevant cells, neuron by neuron" is hardly promising given that the effect of such differences on the behavior of other neurons are dependent on connection strengths that will generally be far from uniform.

- The pervasiveness of inhibition in the nervous system complicates any attempt to define similarity of brain states because it implies that metrics of the weighted sum variety are too simple. If an 'on-center' and an 'off-center' cell, that are generally associated with opposite stimuli, are both more active in brain state A than in brain state B, while in state C the on- and off-differences are opposite, should this lead us to suppose that that from the subjective point of view, A differs more from B than C does? Or the other way round?

The last three objections derive much of their force from the implausibility of the idea that a circumscribed brain system could have phenomenal effects that are strictly independent of its interaction with the neurons with which it communicates: it makes no sense to apply Müller's axioms to anything outside of this circumscribed neural correlate of consciousness. Yet a complete neural model for similarity judgments must provide a causal chain of events underlying decisions of the form "which of these two colors is more like the reference color?", culminating in the binary neuromuscular signal by which the subject indicates his choice. So even if the model does postulate a circumscribed neural system whose activity stands in a one to one relation with phenomenal visual experience, differences in that activity can only be the starting point in accounting for discriminative judgments.

A functional perspective can deal with the difficulties raised, if only by bypassing them initially. Perception is subject to random variation. Experimentally, according to a generalization known as Crozier's Law (Le Grand 1957; Knoblauch and Maloney 1996), differences that are judged just noticeable are ones that are detected with equal reliability, and small differences are judged equal if they are equal multiples of this threshold (Whittle 1973; Whittle 2003). This encourages hope for reducing the psycho-neural gap by measuring the random variation in samples of putatively relevant and accessible neurons: behavioral discrimination can be convincingly modeled on that basis (Shadlen, Britten et al. 1996), and an account of similarity might follow. In this perspective the meaning of a difference in firing rate is determined by its relevance to behavior, and to the activity of other cells, not by well-defined psychophysical linking propositions of the Müller sort where particular phenomenal qualities are inherently associated with particular neural subsystems. This view is consistent with an associationist account of similarity, like the one made explicit in Hayek's connectionist manifesto, *The Sensory Order*, conceived in 1920 and elaborated in 1952 (Hayek, 1952, pp. 53 and 61). Hayek outlines a mechanistic scheme in which he tries to account for everything consequential about sensation while abandoning "the 'absolute' qualities of sensation"—with no obvious regret—as "a phantom-problem". He proposes "that the sensory qualities are not in some manner originally attached to, or an original attribute of, the individual physiological impulses, but that the whole of these qualities is determined by the system of connexions by which the impulses can be transmitted from neuron to neuron; that it is thus the position of the individual impulse or group of impulses in the whole system of such connexions which gives it its distinctive quality; that this system of connexions is acquired in the course of the development of the species and the individual by a kind of 'experience' or 'learning'; and that it reproduces therefore at every stage of its development certain relationships existing in the physical environment between the stimuli evoking the impulses…The connexions…are thus the primary phenomenon that creates the mental phenomena."

What do the color blind see…and why?

Does the meaning of neural signals resides in their relation to other parts of a developing neural system (Hayek), or in phenomenal qualities with which they are inherently associated (Müller)? The sensations of the color blind provide an opportunity to investigate this point. Red/green color blindness is nearly always traceable to a straightforward and simple structural change: a swapping of the L cone pigment for an M pigment, or vice versa (Nathans, Merbs et al. 1992) makes the visual system dichromatic rather than trichromatic. But can this model, together with otherwise unchanged neural processing, account for the sensations of the color blind? It turns out that it can not..

Any attempt to compare the experience of different subjects encounters a familiar obstacle. Through learning, people with normal color discrimination will achieve some degree of consensus in their use of color names, even if the phenomenal experiences on which they base their judgments differ greatly. Two approaches are available to surmount this difficulty. First, as Palmer (Palmer 1999) suggests, perhaps we can trust others to distinguish their phenomenally unitary psychological primaries from the other, compound colors where the phenomenal properties of pairs of primaries are mixed. Even if you and Franz both call grass green, the two of you may do so in virtue of phenomenally different experiences. But if you merely agree about what it means to be composite and what it means to be pure or unitary, that is sufficient grounds for you to trust Franz's claims of the form "what I am now experiencing is an unmixed or unitary color". This doesn't require any precarious psycho-neural linking hypotheses, or theories of neural coding, just a shared understanding of what it means to be unmixed as opposed to composite.

This limited trust can be extended not just to normal observers but to the color deficient. Many red/green blind observers insist that they see red (and greens of sufficiently long wavelength) as a pure and unitary yellow. This claim is credible enough if other colors are reported as compound, but perhaps less convincing if all colors are reported as unitary. Fortunately a second approach is available, thanks to the rare appearance of individuals with one trichromatic and one color-blind eye. For any stimulus viewed with the red/green blind eye, a unilaterally color blind subject can select a color viewed with the normal eye to match it. When this is done, the claims of the bilaterally red/green blind are generally supported: for most of the handful of such observers known to science the sensations from the red/green blind eye range from pure yellow to pure blue. But one clear exception has come to light. MacLeod and Lennie (1975) found a case where red and green appeared orange (610 nm, which he considered far redder than pure yellow); saturated blues and violets appeared slightly greenish blue; and spectrally neutral colors appeared gray or greenish. Thus when represented in the color space of normal vision, the locus of perceived colors derived from the red/green blind eye follows an arc rather than a straight line. In its avoidance of purples and its termination in the orange-red, the arc is directed through a set of colors that are environmentally likely given the information available from the red/green blind eye, and avoids those less likely. This suggests a more complex basis for color appearance than a simple deletion of the red/green opponent signal: the colors seen are roughly ones that minimize the average disagreement between the eyes (Alpern, Kitahara et al. 1983).

Neither the curvilinear locus of our observer, nor the yellow-blue locus of most others, is consistent with a model where only the visual pigments are affected in the color deficient eye. The prediction from that model is quite precise and very different (MacLeod and Lennie 1976): someone with a normal nervous system, but with an L/M pigment swap

that makes the pigments in the 'L' and 'M' cones of the red/green blind eye identical, will match all colors viewed by that eye to a set of colors that stimulate the 'L' and 'M' cones of the normal eye in a constant ratio and are differentiated only by the S cones of the normal eye. This is because if there is bilateral symmetry in postreceptoral processing the binocular match must be a match at the photoreceptor level. At that level, the identity of visual pigment between the 'L' and 'M' cones in the red/green blind eye will make their relative stimulation the same for all color stimuli. So the color stimuli selected as matches using the trichromatic eye must also be equal to one another in their excitation of the 'L' and 'M' cones. The predicted locus of the matching colors is called a tritanopic confusion line, because these colors are confused by tritanopes, who lack S cones and have only the normal 'L' and 'M' cones; the tritanopic confusion line through white connects deep violet at short wavelengths to greenish yellow for long wavelengths.[3] Yet no unilaterally red-green blind subject has reported colors that lie along a tritanopic confusion line. It follows that all known unilateral color blind individuals have atypical neural processing of color, in addition to their pigment swap. The neural change is most likely a reorganization elicited by the altered input from the photoreceptors (MacLeod and Lennie 1976). For the cases where reported colors range from blue to yellow, it might be tempting to invoke atrophy of a red-green opponent system, as Byrne and Hilbert assume (present volume). This does not seem very likely, though, since the parvo (red-green) system is thought to be important for visual acuity, and acuity is unimpaired in the red/green blind. Moreover, loss of input from the electrophysiologically documented "red-green" system (which is driven by L and M cones only), with otherwise normal processing, would lead to perception of violet and lime, not blue and yellow, since violet and lime are the colors that fail in normal observers to excite that system. And for the MacLeod and Lennie case, no such simple account suffices, since redness and greenness are present but are a function of S cone input, in a way that serves to reduce the average discrepancy between the perceptions of the two eyes.

A seldom recognized theoretical point is that in principle, even ordinary (bilateral) red/green blind observers may experience both unitary and compound colors, as MacLeod and Lennie's unilateral case did. And by Palmer's argument above, they should be able to recognize and reliably report this. Subjects whose sensations range along a tritanopic line[4] may experience as much as half of the hue circle, including unitary red and yellow, or unitary green and blue as well as phenomenally compound colors. Only one paper has raised, let alone investigated, these interesting possibilities. Nerger and Cicerone (Cicerone, Nagy et al. 1987) found that "red/green blind" protanopes can identify not only a neutral point in the spectrum, but also could locate a wavelength point in the blue that they described as bluish but (uniquely) lacking in greenness or redness, an observation parallel to those of MacLeod and Lennie's subject.

Thus the sensations of the color blind are quite different from the receptor-based prediction. The mapping from receptors to sensations appears to be highly plastic, in ways that are not yet well documented or well understood[5].

---

[3] Although tritanopes are sometimes called blue-blind or blue/yellow blind, they do *not* confuse blue with yellow, since as noted above these are quite distinct in their effects on the 'L' and 'M' cones.

[4] More strictly: subjects whose sensations are those derived by normals from stimuli that lie along a tritanopic line through white. I have generally resisted pedantic formulations like this on the grounds that even my most carefully qualified language (such as this example) will still be open to philosophical criticism.

[5] One little noted fact about unilateral color blinds is startling enough to warrant mention. At least three of the eight or so putatively congenital unilaterally red/green blinds gave no indication of knowing, prior to investigation, that their two eyes were different! Our subject noticed the difference only when it led to

Plasticity of color perception has also been shown experimentally in people with normal color vision. Long-term exposure to saturated red light over many days caused a persistent (though not completely permanent) redward shift of "pure yellow", as if an adaptive control system adjusts the neural red/green null point toward the centroid of recently encountered color stimuli (Neitz, Carroll et al. 2002). McCollough's observation of orientation-contingent color aftereffects (McCollough 1965) suggests a similar recalibration, with the complication that here, spatial and chromatic perceptions interact. It has been proposed (Barlow, 1990), that in the induction of these aftereffects individual neurons change their chromatic signature, so that after viewing a diet of reddish verticals, neurons selectively sensitive to vertical form change their color preference from achromatic to reddish. Neurons do change their stimulus "trigger features" in this way (Kohn and Movshon 2004). By the same token it is likely, though not yet demonstrated, that the perceptual consequences of activity in early cortical neurons can undergo adaptive modification. Cortical plasticity is a problem for explanatory frameworks that attribute particular aspects of color to neurons with definite identities, and associated connectivities, via well-defined psychophysical linking propositions of the Müller sort. This type of scheme has worked well in correlating color with the earliest neural stage of processing, at the photoreceptors, but it may be inappropriate for the more plastic brain. An alternative can be found in connectionist systems like those of Hebb (Hebb 1949) and Hayek (Hayek 1952), where the stimulus requirements and perceptual sequelae of neural activity are modified as connections are strengthened during development by correlated excitation of the input and output neurons. Such a coupling from monocular inputs to binocular cells would naturally lead to a discrepancy-minimizing pattern of connection for at least some unilaterally color blind observers.

Constant Colors: compensated, or constructed?

In tracing the flow of information from its external source through the photoreceptors and subsequent neural representations, we have encountered more and more fragmentation, and less and less straightforward isomorphism with either the external environment or the phenomenal world that mirrors it. We are in danger of getting lost in the maze. Where to find a way out? What we are looking for is a neural representation more closely isomorphic with phenomenal experience, and hence with the three-dimensional external world that the phenomenal world mirrors. This suggests the possibility that the neural maze has a kind of symmetry, so that the jumble of signals in primary visual cortex gives way to representation that is in some sense more orderly at higher levels of the cortex. Many mazes have a symmetrical construction, so that once you are half-way in you can find an onward path to the exit that is a mirror image of the inward one. Perhaps the neural maze is like those? Not much is known from direct neurophysiological observations to support such a conjecture, but visual phenomena give some clues to the kind of organization that

---

binocular rivalry on one occasion, which led us to bring him into the lab for testing. Only following controlled experiments with binocular matching was he slowly persuaded that his two eyes were different. As a biologist, he had considerable experience of viewing things in a monocular microscope. That one can be red/green blind in one eye without noticing goes beyond oddness. It seems to me to meet the requirements for a philosophically useful datum, in that it nibbles at the edges of conceivability; I for one would not have considered it conceivable until I encountered it. Such cases provide an empirical answer to the question: can an observer be unaware of his own qualia? The answer, in at least a limited sense, would appear to be 'yes', at least in congenital cases.

might be involved in bringing high-level neural representations into closer correspondence with perception.

As noted in the introduction, color constancy provides one example of the closer relation of perception to the proximal than to the distal stimulus. How does perception recover colors and lightnesses that depend closely on the reflectances of external objects, while the retinal stimulus varies due to changing illumination? Fundamental though this question is, our ignorance about its neural basis is remarkably complete.

There is strong support for some elements of Land's early retinex (retina +cortex) model (Land and McCann 1971). Signals from the L, M and S photoreceptors are independently subject to rapid local adaptation that takes the form of a nearly reciprocal adjustment of sensitivity (He and Macleod 1998; Lee, Dacey et al. 2003). Because of this, the signals that pass from retina to cortex when a new stimulus replaces a prior one depend mainly on the ratio of the cone excitations, rather than on their absolute values (Enroth-Cugell and Shapley 1973); this has the effect of compensating fairly well for coloration of the illumination on a scene, as Monge (Mollon 2006) and Helmholtz (von Helmholtz 2004) and Cornsweet (Cornsweet 1970) for example pointed out. If there is no new stimulus, vision simply fails: without motion of the retinal image, objects fade to invisibility over a period of several seconds, much as afterimages do. In normal vision the fading is prevented by small involuntary eye movements that modulate the excitations of the photoreceptors lying sufficiently close to each boundary in the visual field by sweeping the boundary to and fro (Desbordes and Rucci 2007). Although these eye movements critically affect the neural representation, they play no role in neurophysiological recordings, where the animal is paralyzed and the test stimuli have to be flashed, and perhaps for this reason, the neural basis of the fading has not been clearly elucidated despite its obviously fundamental importance. One might hope that retinal output signals monitored electrophysiologically using flashed stimuli would have the same dynamics as the subjectively observed fading, but they never do. Instead, each new stimulus gives rise to a brief transient burst of impulses (or else a comparably transient suppression of firing), that lasts only a fraction of a second and is followed by a weak and variable maintained discharge (Marrocco 1972). It is not clear whether the progressive fading of static images occurs because some cortical circuit responds only to change in the retinal input, or whether it is the brief transient signals that sustain vision, by triggering a persisting change of cortical state appropriate to the new stimulus (as happens in other contexts (Ferber, Humphrey et al. 2003). In either case, the signals arriving at the cortex are very different from the stimulus in their dependence on temporal and spatial as well as spectral parameters. In their strong dependence on chromatic or achromatic *contrast*, they differ radically from the "wavelength-dependent" caricature sometimes drawn by authors who wish to highlight the distinctive contribution of cortical processing to color constancy. But they are also far from being isomorphic to perception, resembling instead (from the point of view of a cortical homunculus) something like a line drawing where lines indicate the boundaries of high contrast. In the retinex model, the cortex resurrects brightness and color at each point by integrating these contrast-dependent signals at nearby and more remote boundaries in the visual field. For the neural mechanism of this tricky accomplishment, there are no well supported conjectures. One interesting class of proposal, consistent with a symmetric scheme where sensory analysis is followed by perceptual synthesis, is the idea (Hurlbert 1986) that the reconstruction could be made by an inverse transform. A high level representation potentially isomorphic with the stimulus is relayed back to primary visual cortex by the brain's plentiful feedback connections. If the descending pathway duplicates the spatiotemporal filtering applied to the afferent signals en

route to primary visual cortex; the high level representation that matches the descending signals to the incoming ones is the correct one. This "predictive coding" scheme is one way to close the loop from the stimulus to the perceptual model, allowing the centrally generated model to preserve isomorphism while the intervening data (the afferent signals) have a quite different form and are used only to test and adjust the model.

However it is managed, the integration of local contrast signals is an incomplete model for color constancy. It fails to explain, for instance, how we are able to distinguish a red scene under neutral light from a neutral scene under red light. Our recognition of that distinction appears to exploit subtle cues, based on the statistical distribution of colors in the image (MacLeod 2003; Brainard, Longere et al. 2006). In view of this, the ideas of a passive correction of the early cortical representation to compensate for changing illumination, and the idea of inverse filtering in a feedback loop, are both too simple. Helmholtz's conception that our estimate of the illuminant represents an "unconscious inference" can accommodate a much wider range of processes. But the neural basis of these inferences is still completely obscure, since on this view the signals monitored in current neurophysiological experiments could be as different from the centrally generated model as scientific data are different from models used to account for them. By the same token, the neural embodiment of the perceptual model is not clarified by investigation of the coding of the sensory data.

It has been tempting to think of color as a kind of bedrock of experience, inherent in sense data and requiring no overlay of interpretation. Yet it is clear that the perceptual specification of color and lightness must include a fairly complete model of the scene, including much information about three dimensional shape and arrangement as well as conditions of observation such as the nature and location of sources of illumination. Phenomena such as the Mach card show that with constant sensory input, perception of lightness and color can be abruptly modified by a revised decision about the surface orientations in a bistable stimulus (Bloj, Kersten et al. 1999; Bloj and Hurlbert 2002). The decision to classify an edge as a shadow or a surface property has similar repercussions. And the estimation of the lightness profile on a curved surface is inextricably intertwined with the estimation of its shape. There is thus no good reason to consider color as "given", while three-dimensional scene geometry is constructed: both are constructed! How their construction is implemented in neural machinery is far from clear, but it seems likely that the cortical input serves only to trigger, guide or direct the perceptual construction of color as well as of other aspects of the environment (Yuille and Kersten 2006).

A recently investigated case of vision recovered in adulthood (Fine, Wade et al. 2003) provides a dramatic and instructive contrast with the normally sighted. Uniquely as far as we know, this subject (MM) has "the eye of the artist" in the sense that he accurately matches the retinal illuminances associated with light and shadowed regions in a scene. (This ability is desirable for artists, but never achieved by them: "phenomenal regression to the real object" (Thouless 1931)seems inescapable.) For MM, the visual world is not automatically perceived as a three dimensional arrangement, and the computation of lightness and color is correspondingly simplified. He does, however, compensate fairly well for changes in the intensity or color of the illumination on the scene as a whole; this may be a result of retinal adaptation processes.

If we make a distinction between sensory processes that deliver a progressively diversifying set of signals along an array of ramifying afferent chains, and perceptual processes where centrally generated hypotheses are tested and modified using feedback-based comparison with the incoming stream of sensory data, constancy seems to owe something to both of these broad stages of processing. But even the relevant sensory

processes are only partially elucidated at the neural level as yet, and the perceptual ones hardly at all. Although the schemes described, involving "generative models" or "predictive coding", feature prominently in current speculation about perception, little or no research in the physiology of color vision has been conducted with such possibilities in mind. If the generative scheme is correct, we would expect some cortical signals to represent the discrepancy between perceptual prediction and the retinal input rather than being directly determined by the input. Some beginnings have been made in testing this in conscious humans, for instance in a study of perceptual organization using brain imaging (Murray, Kersten et al. 2002; Murray, Schrater et al. 2004). Transient visual impairment can be induced by transcranial magnetic stimulation, and in some experiments this has been attributed to interference with neural feedback loops (Pascual-Leone and Walsh 2001), but that observation is not clearly supportive of a predictive coding scheme. Testing the proposal with microelectrodes in animals will be difficult.

Concluding summary

As we have seen, the project of correlating color experience with neural events is hampered not only by the irreducible "explanatory gap" of the metaphysicians, with its attendant "hard problem", but by major additional gaps that reflect our limited current knowledge and understanding of neural processing. The neural representations of color revealed by neurophysiology, particularly in the cortex, are untidy, and new discoveries in neuroscience are only making them more complex. I hope my readers will concur with Hardin (Hardin 1999) that even the "easy" problems we have considered are hard enough—and also that they are far harder than is generally acknowledged. If we are ever to discern an isomorphism between brain events and the experience of color, it will surely require further discoveries and conceptual advances that we can't presently anticipate.

Although receptoral processes are fundamental for trichromacy—the first, and easiest, problem encountered—its basis has not been completely elucidated. And as more central processes become relevant, the significance of neural events becomes increasingly obscure. The post-receptoral recoding of color in terms of opponent signals has long been thought to be a basis for the psychological primaries, but the known opponent signals can not serve as such a basis, and could have quite a different functional significance.

The colors seen by the color blind are not well predicted by their light-sensing abnormality alone; in unilateral cases the colors may be selected, through some kind of neural adaptation, to minimize the discrepancy with what is seen by the trichromatic eye. Although color appearance and color similarity are usually modeled with the assumption that neurons are associated with fixed subjective qualities, like pixels in a Cartesian theater, the plasticity of the relation between afferent signals and perception encourages an associative account instead.

The basis of the "unconscious inferences" on which color vision depends in all but the simplest situations remains obscure. Perhaps feedback to primary visual cortex allows a centrally generated perceptual model to be checked and revised in the light of incoming sensory messages. Such schemes give hope for the possibility of some kind of isomorphism between color perception and as yet unknown central neural events, but they discourage the search for a simple correspondence between subcortical or early cortical neural events and perception.

Acknowledgements

The editors and Rolf Kuehni provided useful comments on an earlier draft. Research supported by NIH grant EY01711.

.

Abramov, I. and J. Gordon (1994). "Color appearance: on seeing red--or yellow, or green, or blue." Annu Rev Psychol **45**: 451-85.

Alpern, M., K. Kitahara, et al. (1983). "Perception of colour in unilateral tritanopia." J Physiol **335**: 683-97.

Arend, L. E., Jr., A. Reeves, et al. (1991). "Simultaneous color constancy: paper with diverse Munsell values." J Opt Soc Am A **8**(4): 661-72.

Barbeito, R., D. Levi, et al. (1985). "Stereo-deficients and stereoblinds cannot make utrocular discriminations." Vision Res **25**(9): 1345-8.

Bloj, M. G. and A. C. Hurlbert (2002). "An empirical study of the traditional Mach card effect." Perception **31**(2): 233-46.

Bloj, M. G., D. Kersten, et al. (1999). "Perception of three-dimensional shape influences colour perception through mutual illumination." Nature **402**(6764): 877-9.

Brainard, D. H., P. Longere, et al. (2006). "Bayesian model of human color constancy." J Vis **6**(11): 1267-81.

Brindley, G. S. (1970). Physiology of the Retina and Visual Pathway. London, Edward Arnold.

Cao, D., J. Pokorny, et al. (2005). "Matching rod percepts with cone stimuli." Vision Res **45**(16): 2119-28.

Carroll, J. (2006). "Colour-blindness detective story not so simple." Clin Exp Optom **89**(3): 184-5; author reply 185-6.

Chalmers, D. J. (1996). The conscious mind. New York, Oxford University Press

Cicerone, C. M., A. L. Nagy, et al. (1987). "Equilibrium hue judgements of dichromats." Vision Research **27**(6): 983-91.

Cornsweet, T. N. (1970). Visual perception, Academic Press New York.

Crick, F. and C. Koch (1995). "Are we aware of neural activity in primary visual cortex?" Nature **375**(6527): 121-3.

Dacey, D. M., H. W. Liao, et al. (2005). "Melanopsin-expressing ganglion cells in primate retina signal colour and irradiance and project to the LGN." Nature **433**(7027): 749-54.

De Valois, R. L., N. P. Cottaris, et al. (2000). "Some transformations of color information from lateral geniculate nucleus to striate cortex." Proc Natl Acad Sci U S A **97**(9): 4997-5002.

Derrington, A. M., J. Krauskopf, et al. (1984). "Chromatic mechanisms in lateral geniculate nucleus of macaque." J Physiol **357**: 241-65.

Desbordes, G. and M. Rucci (2007). "A model of the dynamics of retinal activity during natural visual fixation." Vis Neurosci **24**(2): 217-30.

Dkhissi-Benyahya, O., C. Rieux, et al. (2006). "Immunohistochemical evidence of a melanopsin cone in human retina." Invest Ophthalmol Vis Sci **47**(4): 1636-41.

Enroth-Cugell, C. and R. M. Shapley (1973). "Flux, not retinal illumination, is what cat retinal ganglion cells really care about." J Physiol **233**(2): 311-26.

Felleman, D. J. and D. C. Van Essen (1991). "Distributed hierarchical processing in the primate cerebral cortex." Cereb Cortex **1**(1): 1-47.
Ferber, S., G. K. Humphrey, et al. (2003). "The Lateral Occipital Complex Subserves the Perceptual Persistence of Motion-defined Groupings
" Cereb. Cortex **13**(7): 716-721.
Fine, I., A. R. Wade, et al. (2003). "Long-term deprivation affects visual perception and cortex." Nat Neurosci **6**(9): 915-6.
Guler, A. D., J. L. Ecker, et al. (2008). "Melanopsin cells are the principal conduits for rod-cone input to non-image-forming vision." Nature **453**(7191): 102-5.
Hardin, C. L. (1999). Color Quality and Color Structure. Toward a Science of Consciousness III, The Third Tucson Discussions and Debates
S. R. Hameroff, A. W. Kaszniak and D. J. Chalmers, MIT Press.
Hayek, F. A. (1952). The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology, London: Routledge.
He, S. and D. I. Macleod (1998). "Contrast-modulation flicker: dynamics and spatial resolution of the light adaptation process." Vision Res **38**(7): 985-1000.
He, S. and D. I. MacLeod (2001). "Orientation-selective adaptation and tilt after-effect from invisible patterns." Nature **411**(6836): 473-6.
Hebb, D. O. (1949). The Organization of Behavior. New York, John Wiley.
Hendry, S. H. and R. C. Reid (2000). "The koniocellular pathway in primate vision." Annu Rev Neurosci **23**: 127-53.
Hurlbert, A. (1986). "Formal connections between lightness algorithms." J Opt Soc Am A **3**(10): 1684-93.
Knoblauch, K. and L. T. Maloney (1996). "Testing the indeterminacy of linear color mechanisms from color discrimination data." Vision Research **36**(2): 295-306.
Koffka, K. (1935). Principles of Gestalt Psychology New York, Harcourt, Brace.
Kohn, A. and J. A. Movshon (2004). "Adaptation changes the direction tuning of macaque MT neurons." Nat Neurosci **7**(7): 764-72.
Kolb, H., K. A. Linberg, et al. (1992). "Neurons of the human retina: a Golgi study." J Comp Neurol **318**(2): 147-87.
Land, E. H. and J. J. McCann (1971). "Lightness and retinex theory." J Opt Soc Am **61**(1): 1-11.
Le Grand, Y. (1957). Light, Color and Vision London, Chapman and Hall.
Lee, B. B., D. M. Dacey, et al. (2003). "Dynamics of sensitivity regulation in primate outer retina: the horizontal cell network." J Vis **3**(7): 513-26.
Lennie, P., J. Krauskopf, et al. (1990). "Chromatic mechanisms in striate cortex of macaque." J Neurosci **10**(2): 649-69.
Livingstone, M. and D. Hubel (1988). "Segregation of form, color, movement, and depth: anatomy, physiology, and perception." Science **240**(4853): 740-9.
MacLeod, D. I. and P. Lennie (1976). "Red-green blindness confined to one eye." Vision Res **16**(7): 691-702.
MacLeod, D. I. A. (1985). Receptoral constraints on color appearance. Central and Peripheral Mechanisms of Color Vision. D. Ottoson and S. Zeki. London, MacMillan.
MacLeod, D. I. A. (2003). Colour discrimination, colour constancy and natural scene statistics Normal & Defective Colour Vision. J. D. Mollon, J. Pokorny and K. Knoblauch. London, Oxford University Press.

Marrocco, R. T. (1972). "Maintained activity of monkey optic tract fibers and lateral geniculate nucleus cells." Vision Res **12**(6): 1175-81.
McCollough, C. (1965). "Color Adaptation of Edge-Detectors in the Human Visual System." Science **149**(3688): 1115-1116.
Mollon, J. (2006). "Monge: The Verriest lecture, Lyon, July 2005." Vis Neurosci **23**(3-4): 297-309.
Mollon, J. D. (1989). ""Tho' she kneel'd in that place where they grew..." The uses and origins of primate colour vision." J Exp Biol **146**: 21-38.
Murray, S. O., D. Kersten, et al. (2002). "Shape perception reduces activity in human primary visual cortex." Proc Natl Acad Sci U S A **99**(23): 15164-9.
Murray, S. O., P. Schrater, et al. (2004). "Perceptual grouping and the interactions between visual cortical areas." Neural Netw **17**(5-6): 695-705.
Nagy, A. L., D. I. MacLeod, et al. (1981). "Four cone pigments in women heterozygous for color deficiency." J Opt Soc Am **71**(6): 719-22.
Nathans, J., S. L. Merbs, et al. (1992). "Molecular genetics of human visual pigments." Annu Rev Genet **26**: 403-24.
Neitz, J., J. Carroll, et al. (2002). "Color perception is mediated by a plastic neural mechanism that is adjustable in adults." Neuron **35**(4): 783-92.
Palmer, S. E. (1999). "Color, consciousness, and the isomorphism constraint." Behav Brain Sci **22**(6): 923-43; discussion 944-89.
Pascual-Leone, A. and V. Walsh (2001). "Fast backprojections from the motion to the primary visual area necessary for visual awareness." Science **292**(5516): 510-2.
Shadlen, M. N., K. H. Britten, et al. (1996). "A computational analysis of the relationship between neuronal and behavioral responses to visual motion." J Neurosci **16**(4): 1486-510.
Shady, S., D. I. MacLeod, et al. (2004). "Adaptation from invisible flicker." Proc Natl Acad Sci U S A **101**(14): 5170-3.
Stockman, A., D. I. MacLeod, et al. (1993). "Spectral sensitivities of the human cones." J Opt Soc Am A Opt Image Sci Vis **10**(12): 2491-521.
Thouless, R. H. (1931). "Phenomenal regression to the real object. II." British Journal of Psychology **22**(1): 1-30.
von der Twer, T. and D. I. MacLeod (2001). "Optimal nonlinear codes for the perception of natural colours." Network **12**(3): 395-407.
von Helmholtz, H. (2004). "Treatise on Physiological Optics." Perception.
Vul, E. and D. I. MacLeod (2006). "Contingent aftereffects distinguish conscious and preconscious color processing." Nat Neurosci **9**(7): 873-4.
Whittle, P. (1973). "The brightness of coloured flashes on backgrounds of various colours and luminances." Vision Res **13**(3): 621-38.
Whittle, P. (2003). Contrast Colors. Colour Perception: Mind and the Physical World R. Mausfeld and D. Heyer. London Oxford University Press.
Winer, G. A., A. W. Rader, et al. (2003). "Testing different interpretations for the mistaken belief that rays exit the eyes during vision." J Psychol **137**(3): 243-61.
Yuille, A. and D. Kersten (2006). "Vision as Bayesian inference: analysis by synthesis?" Trends Cogn Sci **10**(7): 301-8.
Zeki, S. (1990). "Parallelism and functional specialization in human visual cortex." Cold Spring Harb Symp Quant Biol **55**: 651-61.