Neurophenomenological Constraints and Pushing Back the Subjectivity Barrier

Extended commentary on Steven E. Palmer's
"Color, Consciousness and the Isomorphism Constraint"

Technical Report UT-CS-99-419

Bruce J. MacLennan*

Computer Science Department
University of Tennessee, Knoxville
maclennan@cs.utk.edu

April 20, 1999

Abstract

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

*This report is an extended version of a commentary to appear in *Behavioral and Brain Sciences*. It may be used for any non-profit purpose provided that the source is credited.

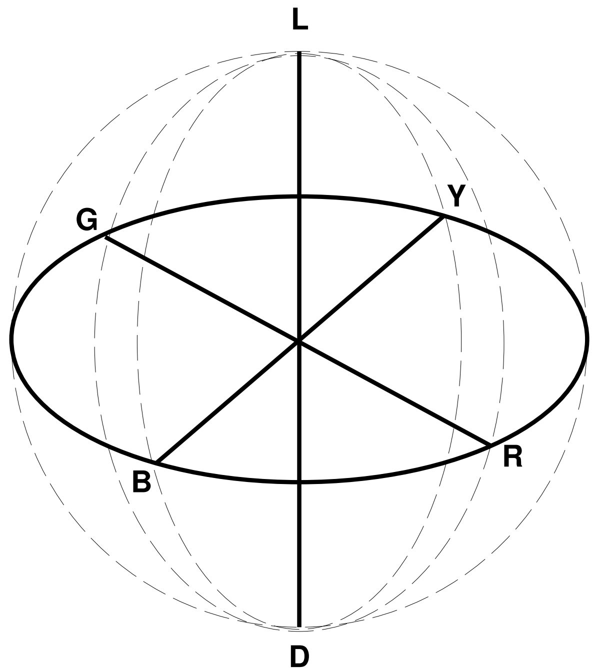

Figure 1: Color Sphere. Capital letters represent phenomenal colors (L/D = light/dark, Y/B = yellow/blue, R/G = red/green).

1 Via negativa

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

Consider first a light/dark (white/black) inversion. The color sphere (Fig. 1) suggests that this is possible, but a phenomenological analysis argues against it, for the light and the dark have different phenomenological structures. As Francis Bacon said, "All colours will agree in the dark"; that is, all hues merge at the bottom of the color sphere. The sphere similarly shows the hues merging at the top, but this seems to be more an artifact of the theory than a phenomenological reality. At best it is a rare experience, such as one might have staring at the sun or into a very bright light. But this reveals another asymmetry of the light/dark axis, for very bright lights are accompanied by pain, but darkness is not. This experience of pain is an integral part of the phenomenology of vision.

I have little to add to the article's treatment of the yellow/blue inversion, except to observe that the "yellow anomaly" (the fact that yellow is inherently lighter than

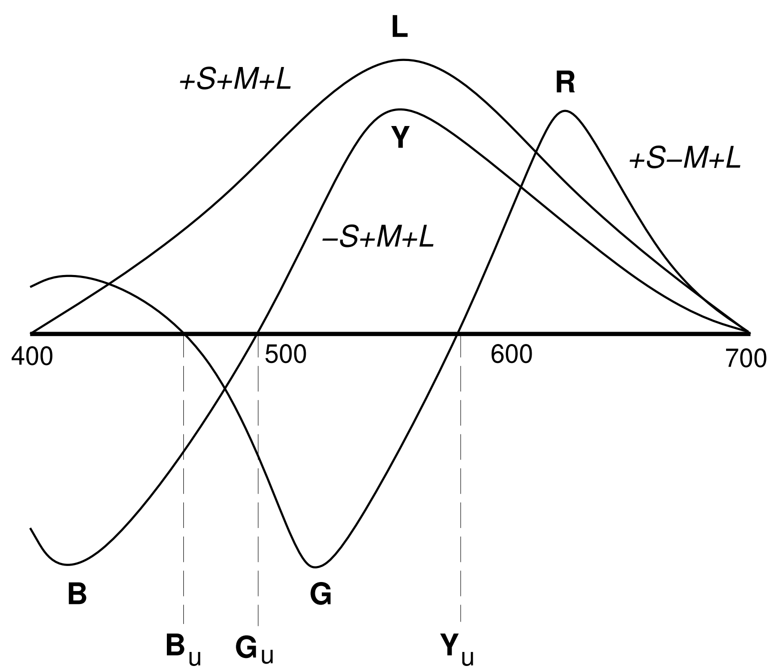

Figure 2: Normal Response Curves of Double Opponent Cells. Capital letters represent phenomenal colors corresponding to peaks in the response curves. Letters with a subscript u represent the three spectral "unique" (unmixed) hues.

the other colors) is predicted and explained by the fact that the response of the yellow channel ($-S + M + L$) has the largest overlap with the light (white) channel ($+S + M + L$) of all the color channels (Fig. 2).[1] (Here neuroscience complements phenomenology.)

Thus, since ancient times (e.g. Aristotle, *De sensu* 442a), phenomenological analyses of color have recognized the similarity between yellow/blue and light/dark, often making them the extremes of a color-series arranged linearly between light and dark. Furthermore, the first two colors in the Berlin and Kay hierarchy are conventionally termed "white" and "black," but are more accurately described as warm-light vs. cool-dark, that is, very much like yellow/blue (Kay & McDaniel 1978).

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

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

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

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

2 Via affirmativa

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

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

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

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

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

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

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

A more problematic abnormality replaces the bimodal channel with a unimodal channel, so that there are two unimodal channels: $-S+M+L$ (and its opposite) and $+S+M-L$ (and its opposite). The problem is that in the worst case there could be complete symmetry between these channels, so they have equal overlaps with the light channels and equal claim to generate the yellow experience (though for different wavelengths). Topologically each of these channels would appear like yellow in its relation to its opposing blue and in the relation of the yellow-blue pair to light-dark. However there would be an additional similarity between each blue and the yellow of the other pair. Here we may have an example of a visual system too alien to permit visualization of the experience, but the topology is clear. In any case the anomaly would be easily detectable, because there would be only two spectral unique hues, as opposed to the three spectral unique hues of normal color vision. (See MacLennan 1999 for more details.)

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

However, it may be argued (Chalmers 1995, 1996; MacLennan 1995, 1996a) that the traditional approach is inadequate for solving the "hard problem," that is, for fully integrating consciousness into the scientific worldview. This is because an adequate account of subjectivity must deal with certain relations between objects (specifically those between subjects and the objects of their consciousness) from a different perspective, from the "inside" of the subjects out to their objects. Such relations cannot be entirely external to the objects. Therefore, we must expand our domain to admit that some objects, at least, have two sides (the outside and the inside, so to speak), one of which is accessible only when the observer *is* the object. Thus we are led to some form of dual-access monism.

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

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

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

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

3 References

1. Chalmers, D. J. (1995) Facing up to the problem of consciousness. *Journal of Consciousness Studies* 2: 200–219.

2. Chalmers, D. J. (1996) *The conscious mind.* Oxford University Press.

3. De Valois, R. L., & De Valois, K. K. (1993) A multi-stage color model. *Vision Research* 33: 1053–1065.

4. von Goethe, J. W. (1840) *Goethe's theory of colours*, transl. C. L. Eastlake. John Murray.

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

6. MacLennan, B. J. (1995) The investigation of consciousness through phenomenology and neuroscience. In *Scale in conscious experience: Is the brain too important to be left to specialists to study?*, ed. J. King & K. H. Pribram. Lawrence Erlbaum. `http://www.cs.utk.edu/~mclennan/papers/ICTPN-tr.ps.Z`

7. MacLennan, B. J. (1996a) The elements of consciousness and their neurodynamical correlates. *Journal of Consciousness Studies* 3: 409–24. Reprinted in *Explaining Consciousness: The Hard Problem*, ed. Jonathan Shear. MIT Press. `http://www.cs.utk.edu/~mclennan/papers/ECNC.ps.Z`

8. MacLennan, B. J. (1996b) Protophenomena and their neurodynamical correlates. University of Tennessee Computer Science Department Technical Report UT-CS-96-331. `http://www.cs.utk.edu/~mclennan/papers/PNC.ps`

9. MacLennan, B. J. (1998) Finding order in our world: The primacy of the concrete in neural representations and the role of invariance in substance reidentification. *Behavioral and Brain Sciences* 21: 78–79. `http://www.cs.utk.edu/~mclennan/papers/FOW-TR.ps.Z`

10. MacLennan, B. J. (1999) The protophenomenal structure of consciousness with especial application to the experience of color (extended version). University of Tennessee Computer Science Department Technical Report UT-CS-99-418. `http://www.cs.utk.edu/~mclennan/papers/PSClong.ps.Z`

11. Palmer, S. E. (in press) Color, consciousness and the isomorphism constraint. *Behavioral and Brain Sciences*.