

COLOR SUBJECTIVISM IS NOT SUPPORTED BY COLOR REDUCTIONISM

Tom Seppalainen

ABSTRACT

If all the participants in the color ontology debate are naturalists with good sciences on their side, how could color subjectivism win? The apparent reason is that subjectivism is supported by the opponent process theory that is a successful neurophysiological reduction of colors. We will argue that the real reason is the unique reductive methodology of the opponent paradigm. We will undermine subjectivism by arguing against the methodology.

1. Introduction

Color ontology has been an area of intense philosophical debate but it seems that the “chromatic zeitgeist” (Hardin, 1993, p. xix) in metaphysics is slowly coming to an end. It is then a good time to think about morals, at least interim ones. Let us start with the fundamental question. Who won the ontology debate? Many philosophers probably lean towards subjectivism (e.g., Hardin, 1993 and Clark, 1993) according to which colors are simply “in the head.” But at the same time many are probably a little unclear on why subjectivists won. This is not surprising since there is a lot to be puzzled about. Even the very idea of a “winner” in this debate is puzzling.

A unique feature of the color ontology debate was that all the serious participants claimed to have a science or sciences of color on their side. In fact this is what made the debate serious, at least in the mind of the participants. Color objectivists (e.g., Hilbert, 1987) claimed that colors reduce to external “physical” properties. The objectivist ontology was based on computational color vision. Subjectivists claimed a part of the

neurophysiology and psychophysics of color to back up their ontology. According to “enactivism” or “interactionism” (e.g., Thompson, 1995), colors are properties of the world that result from animal-environment co-determination. The interactionists’ scientific source comprised all the above together with evolutionary and comparative studies of color vision. We have good reasons to believe that all of these sciences of color are good or at least equally good. They are all based on evidence. If this is true, how did one ontological position come out on top? All three ontologies are after all associated with good sciences which just construe the world of color in ontologically speaking different ways, in the world, in us, and in the process of worldly action, respectively.

In situations where ontological rivals draw from different sciences that are concerned with the same phenomena there can be a clear winner. One possibility is that the winner is less discriminate than any of the ontological alternatives. Instead of a metaphysical “ism” such as subjectivism the winner is a meta-philosophical “ism.” In the color case, this “ism” would be philosophical naturalism. All of the above three ontological theories are naturalistic in pairing ontology intimately with a science or sciences of color. This option does make partial sense of the color case. The color metaphysicians were unanimous about bringing an end to arm-chair speculation. Philosophy of color (and “mind”) was to become philosophy of cognitive science. This attitude is clearly reflected in Thompson’s (1995, xiii) Kantian-type slogan of “cognitive science without philosophy is blind, but philosophy without cognitive science is empty” as well as Hardin’s (1993, xvi) remark that “discussions about color proceeding in ignorance of visual science are intellectually irresponsible.” But naturalism is not a real option for those who believe that subjectivism won the debate. For subjectivism to win, philosophical naturalism would have to favor it instead of the other two ontologies. The second possibility is that the question about the winner of the ontology debate reduces to a question about the best science of color. The ontology debate is won by association with the right science. But if we assume that all the above sciences are good *qua* sciences, this option will not result in *a* winner but, instead, everybody wins. Again, the inclination to favor subjectivism will be without a proper basis.

The color ontology debate is puzzling because if all the participants accept naturalism and all have good natural sciences on their side, how can one manage to win? In abstract terms, the solution to this puzzle is

that one of the sciences at least appears best not in a scientific sense but in a philosophical sense. One of the three sciences (or set of scientific disciplines) is the best one to base an ontological theory on. More specifically, the presence of a winner in the color ontology debate becomes understandable if the methodology of one of the sciences entails an intuitively understandable mind-body relationship. In this paper we will employ this idea to make the success of subjectivism understandable. The summary of our discussion is the following.

Our central analytic goal is to clarify the interdisciplinary methodology of the opponent process paradigm of color vision that the subjectivists based their ontology on. We will show that this paradigm followed an unusual research strategy, a looks-like methodology. We will analyze this methodology in general by discussing some of the vision scientist Davida Teller's (1980 and 1990) writings on the topic and show how it was used in the opponent paradigm. The central feature of this methodology is that the explanatory factors and that which they explain are assumed to resemble one another. The products of this methodology suit ontological conclusions just because of their strong and simplistic reductive nature. After reconstruction we will ask whether the methodology is likely to be successful in a scientific sense. This evaluation proceeds by placing the looks-like methodology in a comparative epistemological context. We will also assess it in light of philosophical discussions of scientific explanation. Our discussion suggests that its success would be highly unlikely.

Our contribution to the color ontology debate is the following. We will both agree and disagree with color subjectivists. We agree that the methodology of the opponent paradigm has been reductive and just in a way that fulfills (maybe) common ontological *desiderata*. However, we disagree on the likelihood of the scientific success of the methodology. We believe it is unlikely to produce a successful reduction. In terms of the ontology debate, our conclusion is that ontological subjectivism does not gain support from neurobiological reductionism as it is represented in the opponent paradigm of color vision. This argument does not inflict color subjectivism at large nor is it intended to. It only concerns the popular reductive argument for it. In order to gain critical perspective on subjectivism itself, we will end with general remarks on the tension between reductive and causal explanations in future cognitive neuroscience.

2. Reduction and Reductionism in Color Science

The opponent process paradigm was introduced to the philosophical community as *the* example of a successful neurobiological reduction of sensations, first by Larry Hardin (1993) and later by Austen Clark (1993). Both claimed that the interdisciplinary efforts of the opponent paradigm were successful at reducing color sensations or qualia to properties of individual neurons. If we ignore all the details and qualifications, the overall structure of their argument for reduction is straightforward. According to this “simple argument,” first, the interdisciplinary opponent paradigm of color vision is reductive and, second, the model of the neurobiological mechanism of color vision is based on good evidence, therefore, color qualia are reduced to neurobiology. This argument led Hardin and Clark to espouse ontological color subjectivism, viz., that colors are “fully in the head.”¹

The color subjectivists argument is *prima facie* persuasive. It sounds like applied science and as such something that at least naturalists should accept. This attitude depends of course on the fact that reductionism is a fairly common epistemological and ontological thesis in many areas of philosophy of science. The argument also suggests that, similarly to other cases of reduction, the task of the philosophers was to clarify the model of reduction that applies to this case (and to convince dualists, functionalists, and other non-believers to get on the bandwagon). The general nature of the writings of color reductionists, however, suggests that the situation is not this straightforward.

The focus of philosophical discussions on reduction has always been on “formal” matters. Philosophers develop models that concern the “logical form” of a reduction. An argument for a reduction requires demonstrating how a particular model applies to a scientific case or cases. For the most part this is not true of the discussion of color reductionism.² The color subjectivists have not shown how any of the popular models of reduction fit the color case nor have they explicated a new model. The description of reduction has been schematic. For example, for Clark a reduction is “a neurophysiological interpretation of the color quality space” (1993, p. 148). This interpretation involves “identifying neurophysiological states and processes that stand in the same patterns of relations as the points of the quality space” (ibid). Instead of a model, the

reductionists have at most provided us with a schematic characterization of the psycho-neural bridge laws. In Clark's view, the bridge laws connect identical or similar "relational" or "structural" properties.

Does the lack of a model entail that color reductionism is philosophically unsatisfactory and not worthy of analysis? There is another possibility. It could imply that color science itself is different and somehow to blame for the philosophical state of affairs. There are good reasons to accept the latter point of view. One reason follows from the fact that reduction is usually taken to mean a relationship between scientific theories. "Theory reductionism" is associated with the idea that reduction is an explanation. The reduced theory is explained because it is deduced from the reducing one with the aid of the bridge principles. A model of theory-reduction will not fit the color case. Both psychology and neuroscience at large are very light on theory but heavy with experiments and demonstrations of various experimental effects instead. This case would require developing a model for a reductive explanation of a phenomenon and not a model of theory-reduction.³

Color science is then at least partly to blame for the nature of the philosophical discussion of color reductionism. As naturalists, the subjectivists must be careful not to employ mistaken conceptual tools such as a wrong model of reduction. But what is there left for philosophers to discuss if they cannot debate formal matters such as the fit between a model of reduction and the scientific case at hand? This question brings us to a distinction that is crucial for the issue of reduction.

"Reduction" is a success word in *two* different senses. The first formal sense is obvious. The relationship between two knowledge representations has to be of a certain sort for the concept to apply. Satisfying formal requirements is not all there is to reduction. Reduction also concerns substantial methodological issues.⁴ Substantial issues of reduction can be of different sorts but all of them concern the epistemic value of reduction. For example, if reduction is an explanation it has to be both reductive and an explanation. If reduction is merely a relationship between two theories, the two theories must be good ones. Clearly the formal and substantial issues are not equivalent. An empirically empty theory could be shown to reduce an empirical fact or well-established theory. It is equally clear that the substantial matters come prior to formal ones. There simply would be no point in attempting to reduce an empirical fact to an empirically empty theory.

Our reason for drawing the distinction between formal and substantial issues of reduction is not to suggest that discussions of formal issues are mere sophistry. Usually the substantial issues can be taken for granted. The paradigmatic cases of reduction involve well-established scientific theories. But can we assume that this is true for color science? Already the nature of the philosophical discussion suggests otherwise. We should expect a lack of concern with formal matters in cases where the substantial issues are still contentious. It is our belief that color reductionism is special for just this reason. It would indeed be premature of the reductive color subjectivists to focus on formal matters under these circumstances.

The reductive color subjectivists in their simple argument make note of the distinction between formal and substantial issues of reduction. The first premise of the simple argument claims that the interdisciplinary opponent paradigm of color vision is reductive. This concerns the nature of the *research strategy* of the opponent paradigm. Let us assume that a reductionist research strategy is one where the theories, laws, and/or empirical facts of one scientific discipline are accounted for by those of another “lower level” or more “basic” science. Of course these higher level “facts” have to be accounted for in a way that is sufficient for reduction. Let us leave the details of what this consists of open. The simple argument moves from a reductive research strategy to a product, a reduction. Let us assume that this product has to be a type of explanation. Reduction refers to the structure of this explanation. The step from a premise concerning a strategy to the conclusion concerning an explanation has to be justified. A research strategy does not automatically result in a successful product, an explanation nor a reductive explanation. Arguing for a reduction on the basis of a mere strategy would conflate between reduction and reductionism. How can the conflation be avoided? A reductive research strategy becomes successful on similar grounds as any other type of research strategy. It must be based on evidence. The second premise of the simple argument addresses this issue. It concerns the evidence for the neurobiological opponent mechanism. It claims that the color case satisfies the substantial criteria of reduction (whatever these might be). This allows the color subjectivist to justifiably move from a strategy to a product.

Unfortunately, the situation is not quite as simple as the above suggests. Distinguishing between formal and substantial issues of

reduction and emphasizing the latter does not reduce an argument for reduction to an issue about evidence *simpliciter*. The argument also runs the risk of triviality. To see this, let us focus on interdisciplinary research projects in which the goal is to account for the empirical facts of a “higher level” science with the empirical facts of a “lower level” science. We cannot assume that such a research strategy is inherently reductionist. If every interdisciplinary area where evidence “flows between” (or probably “flows back and forth” is a more accurate image) a higher and lower level science were reductionist, the term would be largely unnecessary. It would suggest a distinction that is not based on an actual difference.

How does the simple argument avoid the risk of triviality? The evidence for the neurobiological mechanism has to not only be good but it must also be of a type that “goes with” a reductive explanation. The opponent paradigm must display a reductive use of evidence. This means that if we are to discuss the substantial issues of reduction we cannot focus on the use of evidence without relying on *some* ideas about the nature of a reductive product. The second premise of the reductionist’s argument raises then two issues. One concerns the quality of the evidence and the other the reductive character of the evidence. This latter issue reminds us of the importance of the first premise for the truth conditions of the second one.

The above remarks on the concept of “reduction” clarify our agenda. In argumentative terms, we want to critically analyze the simple argument and avoid both triviality and conflation. In order to do this two issues must be discussed. The first concerns the nature of the research strategy of the opponent paradigm. Is it truly reductive? We must search for indicators of a reductive research strategy. But we must do so without a definite model for a reductive product. The color subjectivists are after all naturalists. Still the traditional formal issues of reduction help us here. The methodology of the opponent paradigm should display an attempt to establish “bridge principles” that approximate identities (between properties in the psychological and neurophysiological domains). This argumentative strategy is not based on accepting a model for reduction from elsewhere in philosophy of science. It only rests on the solid intuition that a reductive explanation must somehow differ from a causal explanation. There are good grounds for not requiring that causes and effects share in properties. A similar requirement does apply to reduction.

In addition, if we do not distinguish between the two explanation types, “reduction” becomes an empty concept. This would make the simple argument trivial. Once we have identified a reductive research strategy we can address the second issue. Is it likely to be successful so that the simple argument can avoid the conflation between reduction and reductionism? In other words, is it likely that the product is both a reduction and explanatory? At this stage of color science our answer will be tentative. The best we can do is to evaluate the research strategy and the ensuing explanatory structure in a comparative context.

3. Reductionism and the Looks-like Methodology of Cognitive Neuroscience

The scientific source for the reductive color subjectivist is interdisciplinary. It consists of psychophysics or “perceptual psychology” and neurophysiology of color. We will follow common usage and locate this interdisciplinary effort in the cognitive neurosciences (although many would deny that color vision is properly called labeled cognitive). What are the general methodological features of this interdisciplinary research? How does evidence flow between them?

Many cognitive neuroscientists and philosophers of psychology emphasize the role of psychophysics for neurophysiological inquiries. The color scientists Lennie and D’zamura (1988, p. 391) give a lucid expression of this:

Without clearly articulated ideas about the probable properties of the underlying mechanism, physiologists will find it very hard to piece together an authoritative account of their operation. In short, the physiologists need strong guidance as to the kind of mechanism they should expect to find.

According to Lennie and D’zamura, the neuroscientist needs advice on the probable theoretical properties of the neurophysiological mechanism. The source of this information is the psychophysicist. What type of a methodology guides this interdisciplinary effort? We must be able to explicate some of the neurophysiologists’ strategies for using psychophysicists’ information on perceptual “effects” in order to judge

the reductive character of their research programs.

David Teller's insightful ideas (e.g., 1980 and 1990) can be used to shed light on this methodological question. Teller discusses the issue in terms of "locus questions." "What is the site of dark adaptation?" and "Where do Mach bands occur?" are two of her examples of locus questions. Both dark adaptation and Mach bands are perceptually perspicuous mental phenomena. In the former, a stimulus that first appears dark starts to appear lighter over time. The latter involves stimuli with intensity ramps that enhance the brightness appearance of adjacent areas in a perceptually striking manner.

The question about the brain site or locus of the above "psychophysical" phenomena is a first step towards our concerns. An answer to a locus question requires connecting psychophysics with neuroscience. It will require providing psychophysical evidence for a neurophysiological mechanism. Still, locus questions sound fairly indiscriminate. They sound like question about the localization of mental functions. However, Teller's questions demand a fairly specific answer. A brain area lighting up in a PET scan, for example, is not the appropriate type of answer. According to Teller (1980, p. 156), answering the locus question for Mach bands, for example, is "making a guess about how subjective brightness is represented in the physiological system." Locus questions are about the *neural representations of perceptual phenomena*. The representation is at a brain site or locus.

Let us think for a moment about how Teller in posing locus questions conceives perception and its underlying mechanism. She conceptualizes a color experience, for example, as the result of a complex chain of events. Part of the chain is internal to the visual system. The internal part connects the conscious color experience with the light hitting the eye. The internal chain consists of intermediate events or stages of visual processing. For theoretical purposes, it is thought of as a series of mappings. Teller's questions combine the issue of a neural representation with that of a brain site. The answer to the locus of a color experience is not the full chain of events or the full set of maps. This entails that each stage of processing is not of equal importance. The site that harbors the representation of the color experience is far more important than the rest. In other words, the question assumes that a privileged mapping exists between one stage of neural processing and the perceptual response or

color experience.

Teller's view of visual processing can be also described in the information-processing metaphor. Information flows through the color vision system. It undergoes various transformations along the way. The transformations are described as mappings. The assumption behind locus questions is that some of the information present at different stages of processing appears in consciousness. It is assumed that some stages have more of the information that is present in consciousness than others. The locus is the one that has most of the information that is present in consciousness. It is the stage that is in the "clearest" or "strongest" mapping-relation to the conscious response:

What are the actual methods used to answer locus questions? Teller discusses several ways to answer them. According to one, the locus is the stage of the neurophysiological system where the physiological signals caused by the stimulus start looking more like the perceptual response than the stimulus (1980, p. 155). The strongest version poses the question, "at what level of the visual system do the signals look the most like the response" (ibid.). Both answers employ the relationship of "looking alike." According to Teller's methodology, the critical factor in answering a locus question is the similarity between the neural responses and the perceptual responses. It is a "looks-like methodology." For the purposes of our paper, we will adopt the latter version. According to the looks-like methodology, *the locus of a perceptual response is that stage or level of processing at which the neurophysiological response looks most like the perceptual response.*

The looks-like methodology requires comparing mental and neurophysiological responses and assessing their similarity. This should not be taken too literally. It does not mean that the neurophysiological signal for a color experience, for example, has to be colored. In general, "looking alike" should not be confused with ordinary examples of resemblance such as the way in which a photograph of a scene and our perception of it resemble one another. According to Teller, resemblance pertains to "abstract properties" (1980, p. 156). Her clearest example concerns the resemblance between a graph representing psychophysical data and a graph representing physiological data. If the two graphs can be manipulated so that their shapes look alike then the physiological phenomenon can be assumed to be the critical causal factor for the psychophysical phenomenon (Teller, 1990, p. 13). This is tantamount to

identifying the neural representation of the perceptual response.

The looks-like methodology provides a way in which the psychophysicists' knowledge of perceptual responses can be put to use by the neuroscientists. If theoretically important properties of the neural mechanism resemble mental properties, the neuroscientist can make use of psychophysical facts in the way described by Lennie and D'Zmura above. A locus question can be investigated by first settling on the criteria for the neurophysiological representation through psychophysical means. Once the psychophysical data has been analyzed, the neuroscientist can start searching for the locus and the representation by interpreting neural responses in light of the looks-like methodology. This is achieved by judging the similarity of the two responses. Once a match is found, the neuroscientist will have the critical neural property of the underlying mechanism.

Let us now describe the general nature of this research strategy. Teller thinks that something of a "peculiar slip of the mind" occurs when cognitive neuroscientists ask locus questions. The questions seem to imply that "causes of the effect reside in some parts of the physiological chain more than in other parts" (1980, p. 152). It is not immediately clear whether this is a "slip of the mind" and not just a "slip of the pen." After all, the idea might be that some stages or levels of processing are more important than others for the causation of the experience, although all of them are necessary. Some causal factors might turn out to be weightier than others and in this sense the cause can "reside more" in some factors than in others. Teller's wording sounds odd but there might be nothing unwarranted about this view of causation.

Regardless of the above remarks, there is something to Teller's self-criticism. In Teller's description, the looks-like methodology runs together two concepts, the neural cause of a mental response and the neural representation of that response. The "slip of the mind" occurs here. Usually we do not equate the relationships of causation with the relationship of representation. At least, we do not think that a representation is just any item in a causal chain. The problems that plague causal accounts of mental representation indicate this much. Our methods of causal analysis also indicate that we should not equate the two relations. Using the looks-like methodology in causal inference would entail that the brain locus is a curious cause. This cause would resemble its effect. At least empiricists have always been suspicious of such

methods of causal inference (see section 5 below).

It is not difficult to avoid Teller's "slip of the mind." Why should the concept of a neural representation of a mental response be associated with a causal research strategy in the first place? After all, the tension here results from coupling the idea of a chain of largely "democratic" causal factors with the idea of a clearly privileged representational factor. Why not think of the research strategy as a reductionist one? A couple of reasons suggest we should indeed do so. First, the neural representation is at a site. Second, the representation represents through a similarity relation. Combining the two ideas entails that at least part of the mental response really is at the neural locus, at the representational site. Of course similarity is not the same as an identity so the methodology is not reductionist in an orthodox sense. Nonetheless, a similarity or partial identity between aspects of an *explanandum* and an *explanans* coheres better with a reductive explanation than a causal one. Given the above considerations, we have reasonable grounds to conclude that cognitive neurosciences that employ a looks-like methodology follow a reductive research strategy. They espouse reductionism

4. The Reductionist Looks-like Methodology of the Opponent Process Paradigm

The simple argument for reduction is based on the opponent process paradigm of color vision. In order to discuss the reductive nature of this paradigm we must analyze the role of psychophysics for the neurophysiological inquiries. We will use the conclusions of the previous section to frame our questions: Have the researchers employed a looks-like methodology? Have the neurophysiologists employed resemblance-based reasoning on psychophysical premises to identify probable properties of the color vision mechanism? Affirmative answers to these two questions will indicate a reductionist research paradigm.

Hurvich and Jameson (e.g., 1957) opponent process model is central to the interdisciplinary opponent paradigm. According to this "hypothetical" psychophysical model, the human color vision system consists of three color channels, a red-green, a blue-yellow, and a white-black channel. The channels are directly responsible for the type of color sensations in their names. They secrete hues and achromatic sensations.

The model was developed primarily to account for the following two psychophysical effects or perceptual responses (see, e.g., Hurvich, 1981, p. 11).

The first phenomenon concerns the building blocks of human color appearance or "primary colors." Scientists of the opponent paradigm believe in six primary colors, red, green, blue, yellow, white, and black. Each perceivable color is either a primary color or a mixture of some of the six. The second phenomenon concerns the relationships among the primary colors. The central effect is the "opponency" of red and green on the one hand and blue and yellow on the other. These primaries do not mix in a perceptually transparent manner. If two stimuli that individually appear red and green are physically mixed, the mixture will not appear reddish-green or greenish-red. The same is true for blue and yellow. Color opponency is often described as an "antagonism" or as the "mutual exclusiveness" of the two hues in the two pairs. Mixtures of other primaries are different. For example, humans can readily distinguish red and yellow from an orange stimulus. Contrary to the opponent hues, other primary colors form "binary" mixtures such as orange.

Hurvich and Jameson's model accounts nicely for the above two phenomena. The hue channels secrete only primary colors and both of the hue channels display opponent processing. A light stimulus can only result in one of the mutually exclusive hue responses of a channel at any given time. The red-green channel, for example, can only respond with either red or green. This accounts for color opponency. The phenomenon of binary colors is explained by the fact that the effects of the two separate hue channels are combined in a perceptually transparent way. Orange results from the responses of both hue channels with red and yellow being the "winners" of their respective channels.

Let us turn to the neurophysiology of color. The goal of a neural mechanism for color appearance is to account for the above two perceptual effects (among other things). However, instead of employing hypothetical constructs such as "color channels" and "opponent processes" a neural model must consist of "real" units and responses. Neurophysiologists attempt to achieve this by postulating neural cells that "code for" the primary colors. The backbone of the neurobiological color mechanism is a classification of neural cells into types on the basis of the primary colors they "code for." Our task is to investigate the

methodological underpinnings of this classification. We will focus on the classification principles employed by Russell De Valois and his collaborators (De Valois, 1965 and De Valois, Abramov, and Jacobs, 1966). The results of their research form the neurophysiological counterpart of Hurvich and Jameson's psychophysical opponent model. They are also the empirical basis of the color subjectivists' argument (Hardin, 1993, pp. 54-55 and Clark, 1993, pp. 149-153).

Let us recall the looks-like methodology. It directs the scientist's attention to perceptual and neural responses and requires her to compare the two. The responses of neurons are investigated by microelectrodes that measure the electrical discharge of cells. The stimulus of interest is placed in the cell's receptive field. The cell's electrical discharge is measured usually before, during, and after the termination of the stimulation. The results are expressed in terms of firing rates or spike frequencies. How can differences in firing rate be conceptualized in a way that makes a comparison between them and perceptual responses possible?

One aspect of the comparison might seem fairly straightforward. In color perception, humans respond selectively to light wavelength. Cells that code for color have to respond in a similar fashion. They have to have a similar "spectral response profile." Of course constructing such profiles requires deciding first what a cell response is. Only then can one correlate a response with light wavelength, construct a profile of the cell responses as a function of wavelength, and, finally, compare this with the human responses. It is far from obvious how a cell's response should be conceptualized. In addition, humans do not *just* discriminate among light stimuli. The perceptual effects we discussed earlier remind us of this. The perceptual color world comes in primary colors. It also displays curious relationships among the primaries such as opponency. These perceptual responses can be useful if the researcher follows a looks-like methodology. They can be used to discover and interpret neural responses. Did the looks-like methodology play a role in De Valois' conceptualization of neural responses?

Neurons respond when certain types of stimuli are placed in their receptive fields. Electrical discharge, however, is present even in the absence of stimulation. This spontaneous activity is used to establish a base level firing rate. A cell response is then defined as a difference in firing rate from the base rate. How to proceed further in conceptualizing

cell responses? Prior to De Valois' work, it was common to employ Hartline's (1938) "on-off principle." According to Hartline, cells that respond or display firing rates different from the base rate when a light is flashed in their receptive field are "on-cells." Cells that fire when the light is turned off are "off-cells."

De Valois first employed Hartline's principle himself. Soon he realized that it was based on a tacit assumption:

Implicit in such a scheme (the "on-off" classification of a cell response) is the assumption that only excitation is of importance; the decreases in firing shown by these cells are not alluded to in the classification. It would seem more appropriate to classify cells on the basis of the kind of responses they give to stimulus we find no ground for the assumption that only increases of firing are of importance. People have ignored the fact that these cells are responding during the stimulus, but that their response is inhibition (1965, p. 152).

De Valois rightly saw that the "on-off" principle was based solely on *increases* in firing rate. Hartline distinguished between cells on the basis of the time of the occurrence of an increase in firing rate (with respect to the presence of the stimulus). His tacit assumption was that only excitation is a "real" response. De Valois found this assumption unwarranted. He assumed, in contrast, that *inhibition* or a decrease in firing rate was also a response. Correspondingly, De Valois distinguished three different neural responses, inhibition, excitation, and base rate. The threefold distinction of cell responses creates the backbone of the neurophysiological color mechanism. It consists of three categorically different types of cells. Color opponent or spectrally opponent cells form the hue coding mechanism. They respond with excitation to some lights and with inhibition to others. "Excitator" cells respond with excitation to all lights. "Inhibitor" cells respond with inhibition to all lights. The two latter types code for achromatic colors.

De Valois' decision to replace the "on-off" principle by the "inhibition-excitation" principle was a re-conceptualization of the very concept of a neural response. What justifies and motivates this theoretically critical assumption? The reason behind it is revealed in De Valois' "basic assumption" (*ibid.*, p. 159) about the spectrally opponent hue coded cells:

[T]he R+G- cell is signaling red with an increase in activity and green with a decrease in activity, i.e. that the animal will see red when this cell fires rapidly and green when the cell is inhibited.

In the above quote, De Valois describes one of the color-coded cell types of the hue mechanism. It displays opponent responses. Red lights excite it and green lights inhibit it. According to De Valois' assumption, cells such as R+G- are directly responsible for the primary hues. In the R+G-cell, inhibition is responsible for green and excitation for its opponent hue red. The cell "codes for" red and green by its two different response types. The full hue mechanism consists of the many color coded cells that fall into one of four types (two red-green and two blue-yellow types). But why think that opponent cells code for hues in the first place? Thinking about this question reveals the importance of the "inhibition-excitation" principle.

Recall that perceptual color responses are opponent or mutually exclusive. Neither red and green nor blue and yellow occur simultaneously. Hurvich and Jameson postulate hypothetical opponent processes to account for this perceptual fact. De Valois' "basic assumption" accomplishes the same but in terms of "real" neural units and responses. Clearly neural inhibition cannot occur simultaneously with neural excitation so the two neural responses are mutually exclusive. Opponent responses or processing occurs also in each cell of the neural hue mechanism. But to "discover" opponent responses, one must let go of Hartline's principle. "On" and "off" responses are not opponent (both refer to an increase in firing) whereas inhibition and excitation are. De Valois' basic assumption and the "inhibition-excitation" principle that underwrites this assumption are based on the similarity between neurophysiological and psychological responses. De Valois reasoning follows the looks-like methodology.

De Valois' re-conceptualization of Hartline's "on-off" principle should be seen in the context of the looks-like methodology. The case is also a vivid example of the situation described by Lennie and Dzmura. According to them, the neuroscientist needs help from the psychophysicist in her task of identifying the "probable properties" of neural mechanisms. The perceptual effect of opponency discovered by psychophysicists was critical for De Valois in making sense of cell responses. He did not treat

a decrease in firing rate as irrelevant neural fluctuation or “noise.” Instead, he conceived it as a real neural response type. The reason was that it was of the correct type in light of the perceptual response of opponency. Inhibition could be thought of as opponent to excitation. The perceptual effect becomes evidence for the functional or theoretical relevance of both inhibition and excitation.

What about reductionism in the opponent paradigm? It clearly displays the reductionist nature of the looks-like methodology. This methodology requires us to search for one critical explanatory factor, the unit with a neural representation or code. The code is settled on the basis of resemblance. In the opponent paradigm, the identification and confirmation of the neurophysiological determinants of color responses are based on considerations of resemblance between the two. Correspondingly, the methodology assumes that for an adequate neurobiological explanation, the neuron’s responses have to resemble the perceptual responses. In his basic assumption, De Valois’ clearly expresses his commitment to this explanatory model. The research strategy of the opponent paradigm is reductionist. The first premise of the color subjectivist’s simple argument is true.

At this point we can revisit the initial puzzle about the winner of the ontology debate. The looks-like methodology makes the inclination to favor subjectivism understandable. The methodology has intuitive ontological appeal. For ontological concerns, it is hard to even imagine what could be a better reason for privileging one over the many other determinants of color than the discovery of identical (relational) properties between that one determinant and color. Although initially puzzling, under these circumstances it would seem intuitively appropriate to base the color ontology fully on neurobiological factors despite the scientific relevance of the factors postulated by the other good sciences of color. Having said all that, we must not forget the second premise of the simple argument that concerns the evidence for the neurobiological mechanism. Has the color subjectivist shown that color qualia are *successfully* reduced? To answer this question we must turn to an epistemological evaluation of the opponent paradigm.

5. Causes of Concern About Color Reductionism

There is no knockdown argument against the reduction of colors at the current stage of color science. No one has shown, for example, that creatures with only inhibitory responses at the opponent cells do not perceive red and yellow but only green and blue.⁵ The empirical verdict is still out. Still, we can gain critical perspective on neurobiological color reductionism by placing the looks-like methodology itself in a comparative epistemological context.⁶

Let us explore the looks-like methodology in the context of genetics. For the sake of argument, let us assume that genetic determinism is not a mere strawman. In other words, we will assume that some people believe that DNA molecules or genes “code for” traits such as illnesses or behavioral patterns. A common criticism of this form of genetic determinism is that it resembles the old “preformationist” explanations in biology. Lewontin (e.g., 2000), for example, argues that the explanatory pattern of genetic determinism is analogous to the 17th century image in which a child is depicted as fully formed inside the sperm cell. Development is merely an unfolding of the “spermal inner child.” The child (the effect) is represented in miniature form inside the sperm (one of its explanatory causes). Although genetic determinists explanations of development are certainly more complicated, they still assume that the final outcome or trait is fully “in the gene.” The difference between the two is only in the nature of the representational or “coding” relationship. Today’s relationships are no longer pictorial. Instead, development is seen as an unfolding of the “information” that exists in the molecule in “coded form.” The picturing-relation has been replaced by a coded information-relation. Nonetheless, according to Lewontin and other critics such as Oyama (e.g., 2000a), genetic determinism still assumes that the outcome or trait is fully represented in the gene. For our concerns, it is important to note that the preformationist view is an extreme example of the use of a looks-like methodology. The child resembles the spermal cause to the extent that only the proportions of the two differ. If we believe the above critics in their claim that genetic determinists still deposit the outcome fully into the gene, we can maintain that it too espouses a looks-like methodology of sorts.

Lewontin and Oyama are not critical of genes as explanatory factors. They are against the idea of genes coding for traits. Codes and coding are

misleading metaphors since they imply that representations of traits exist in genes. This is a mistake regardless of whether we think of representations as “coded instructions” (“imperative representations”) as genetic determinists do or as descriptions of how things are (“indicative representations”) as the preformationists of the old did. There are two main reasons for this conclusion. First, traits are an effect of a multitude of different factors. The DNA molecule is only one factor among many other internal as well as external or environmental ones. It is epistemologically illegitimate as well as misleading to privilege one factor as the code. On grounds of parity, one could just as well conclude that environments code for traits.⁷

In addition to the above empirical point about biological dependencies, the critics raise an issue about the nature of explanation. By bringing up the preformationist image, Lewontin is critical of the model of explanation assumed by genetic determinists. A “spermal” duplicate of the child is really no explanation at all. The lack of agreement among philosophers of science on the nature of scientific explanation does not touch this point. We should all accept that a minimal requirement of an explanation is that the *explanandum* is explained by something other than itself. The preformationist image violates the minimal criterion. Oyama (e.g., 2000, p. 73 and p. 163) attacks genetic determinists explicitly on these grounds. For her, factors such as codes and programs that are deposited in the gene explain nothing but appear to explain everything. According to Oyama, it is a sign of a pseudo-explanation that the phenomenon to be explained is contained in the alleged explanatory factor.⁸ Genetic determinism really qualifies as a pseudoscience since it purports to but fails to explain.

The above two arguments against genetic determinism can be used to gain critical perspective on the looks-like methodology of the opponent paradigm. Let us think of color experiences as analogous to traits. In the opponent paradigm, neural cells are assumed to code for them. The genetic determinists construal of genetic coding fails on empirical grounds because traits depend on many other factors in addition to the DNA molecules. Does this empirical point apply to the color case? It is difficult to give a conclusive empirical answer at the present time. Still, there are no good reasons to believe that a color experience would depend solely on a factor, the neuron with a color code. It is far more reasonable to think that a color experience depends on a very complex “pathway,”

“chain,” or a “web” of interactions. In all likelihood there are as many cognitive neuroscientists that accept the latter as there are geneticists who accept the idea that traits depend on many factors. Teller is clearly one of these since she does assume a “chain” of causes (e.g., 1980, p. 152). But as we argued above, combining the idea of a “chain of factors” not to mention a “web of interactions” with the idea of a locus that carries a code is contradictory. The “slip of the mind” of the looks-like methodologist that Teller talked about refers to this contradiction. We have to let go of the idea of one special factor at a site and replace it with the idea of many different factors at many sites to avoid the contradiction. We have good reasons to believe that the empirical “multi-factorial” moral from genetics applies to the neuroscience of color. This premise argues against the looks-like methodology and, hence, against the empirical success of the neurobiological reductionism of the opponent paradigm.

The second moral from the case of genetic determinism also applies to the color case. Genetic codes are a pseudo-explanation if they merely deposit the phenomenon to be explained in one explanatory factor. The looks-like methodology assumes the opposite model of explanation. According to it, the more similar the responses of neurons and perceptual responses are, the better the explanation. The explanations that the color subjectivists put forth come close to violating the minimal requirement for an explanation. The *explanandum* and *explanans*, if not fully identical, at least share properties such as opponency.

One might object this argument by simply claiming that coming close to violating a requirement is not a violation, only close to it. This is of course true. Structural similarities or identities between relational properties such as opponency are a far cry from identity *period*. However, we should note that it is just these shared relational properties that make one factor explanatory of the other. In addition, the looks-like methodology does entail that the closer the resemblance between the *explanans* and *explanandum* is, the stronger the explanation. A logical consequence of the methodology is that a full identity between the *explanans* and the *explanandum* would be an ideal form of explanation. The case of genetics suggests that this would be the ultimate form of a pseudo-explanation. In other words, a neural color code is meant to be explanatory in a way in which genetic codes are not explanatory. Why should we assume that it is explanatory to deposit the perceptual response

in coded form in a neuron if it is not explanatory to deposit a trait in coded form in a gene? The opponent paradigm makes a virtue out of a principle which elsewhere in biology is considered a vice. It is difficult to think of good reasons to justify these opposite views on explanation in biology.

On the basis of the comparison between genetics and cognitive neuroscience we should be skeptical about the prospects of success for the *type* of reductionism true of the color case. The alleged reduction is an outcome of a simplistic and questionable methodology. The features that make it questionable are just the ones scientists and philosophers are worried about in at least the general public's understanding of genetics. Given the above discussion of explanation, it is not surprising that many examples from mainstream pseudoscience follow a looks-like methodology. Let us end our critical analysis by describing a few.

Adolf Grünbaum (e.g., 1984 and 1993) has shown conclusively that Freud's reasoning about pathological factors was based on a looks-like methodology. For example, Freud traced the causation of hysterical vomiting to "an experience which justifiably produced a high amount of disgust, for instance, the sight of a decomposing body" (quoted in Grünbaum, 1993, p. 136). Grünbaum describes Freud's method as causal inference on the basis of "thematic affinity." This is a looks-like methodology of sorts. For Grünbaum, this methodology commits the "thematic affinity fallacy" (1984, p. 55). In judging the inference a fallacy, Grünbaum joins company with many other empiricists who are skeptical about a looks-like methodology.⁹

There are many other examples. Historical narratives are often structured in terms of significant individuals. Such "great-man explanations" follow a looks-like methodology of historiography. Historical events are planned and therefore the outcomes or effects will resemble their causes. Otherwise the events would not have happened as planned by the significant individual. Conspiracy theorists attempt to explain unfortunate societal events by invoking "minds" with intentions. Events are explained by representations in the mind of intelligent beings instead of being explained by many different events coming together without a representation of the outcome in any one of them. Given the bad company that the looks-like methodology keeps and the simplistic view of the determinants of perceptual events assumed by it, we would be wise not to believe in the reductive products it produces in the color

case.

6. Concluding Remarks.

The reductive color subjectivists have applied a scientific paradigm to solve a problem in the ontology of the mind. We have argued against subjectivism on methodological grounds and showed why we should not apply this science to solve our philosophical problem. In doing so, we have accepted the subjectivists' naturalism but beaten them at their own game. For concluding purposes, let us speculate on future naturalistic solutions to ontological problems of the mind in the light of the case we analyzed.

The opponent paradigm follows a looks-like methodology. This makes it a unique science. At the same time this makes it metaphysically attractive. A cognitive neuroscience based on this methodology entails a mind-body relationship that is far more intimate than the usual "operational" relations of science such as "correlation." This is why the color reductionists thought it was possible to win the ontology debate by simply picking a mechanism from color science. They would have never bothered with the opponent paradigm if it had produced only operational notions. This is not surprising since, even as naturalists, they accepted the philosophical nature of the mind-body problem. They, like the rest of us, wanted to describe and "understand" the mind-body relationship. Thinking of the looks-like methodology as a causal one does not change this fact. In this case, the looks-like methodology would entail that causes and effects resemble each other. This would make it a unique causal methodology and, again, one that delivers intimate and intuitively understandable relations between the mind and the brain. Such notions capture our ontological imagination and match simple-minded ontological intuitions.

We can draw some conclusions about the future of naturalistic solutions in the ontology of mind by thinking about the epistemological future of the looks-like methodology. It is clearly an unstable scientific methodology. The tension between the idea of a web or chain of causes on the one hand and a special code or representational site on the other hand indicates this. The methodology straddles the fence between reduction and causation. For a causal explanation to incorporate the

desiderata of reduction would entail that causes and effects share properties. This makes the causation of many if not all events too simplistic to be plausible. The scientists will inevitably choose a less simplistic methodology of causation. Of course they will do so on the basis of the available experimental methods and not on principled grounds.¹⁰

In the future, the neuroscientists will continue chipping away at the determinants of our mentality at any level of brain hierarchy that they can get their epistemic hooks on. The neuroscientists will not straddle the fence between causation and reduction forever. Eventually they will let go of the looks-like methodology. This spells doom for the reductive subjectivist. Once the looks-like methodology goes, the naturalistic metaphysician must let go of the hope of a simple and intuitively understandable reductive relationship. This is not a surprise for true naturalists. They know that a naturalistic ontology is as unstable as the epistemological status of the science it applies. For future reductive naturalists, the developments in neuroscience will require elaborating a real model of a reductive mechanism. Of course color subjectivism might still prevail. Then again, if genetics is an indicator, it might not. The causal chain might have to be extended "beyond the head." In any case, even if colors are stuck "in the head," they won't be there for the wrong reasons.

Portland State University

NOTES

1. For these subjectivists, the philosophical significance of the case goes well beyond colors themselves. Hardin espouses metaphysical reductionism. He uses today's color science to argue against all irreducible mental entities such as sense-data. Clark advocates general epistemological reductionism. The color case shows that neuroscience can explain qualia. The mind is not plagued by a special "explanatory gap." A successful explanatory strategy already exists in color science and it is reductive. Given these metaphysical and epistemological implications, a lot rides on the simple argument
2. Paul Churchland (e.g., 1985) is an exception since he has developed a detailed model of reduction that he argues is applicable to sensory neuroscience. However, Churchland's color reductionism suffers from a

- lack of empirical detail. His argument for color reductionism is on an in principle level and does not concern the opponent paradigm at all. Hardin and Clark's subjectivism does not then gain any support from Churchland's argument and so it will not be addressed in this paper.
3. The lack of a model of reduction implies that debates on reductionism in psychology that focus exclusively on the nature and existence of bridge principles are without proper context. Philosophers often object mind-body reduction on grounds of "multiple realization" or other similar notions that indicate the lack of bridge principles of the appropriate type. But this issue has been carried over directly from debates in theory-reductionism. It need not be as important in the context of reductive explanations of phenomena. At least, focusing on the nature of bridge principles without the context of a full-blown model of reduction results in unprincipled debates. This is clearly visible in the color case and the recent debate between Dedrick (1996) and Clark (1996). The backdrop for this debate was the subjectivists' criticism of color objectivists (e.g., Hilbert, 1987) who claimed that colors reduce to external "physical" properties. Subjectivists argued against objectivists on grounds of the many-one problem. According to the subjectivists, the phenomenon of metamers refutes objectivism because many physically different stimuli appear the same color. According to Dedrick, the many-one problem applies also to reductive subjectivism. Many different neurophysiological states of the opponent color channels correspond to the same color. On grounds of parity, the many-one problem should also refute color subjectivism. Clark dismissed Dedrick's concerns on more complex argumentative grounds but his central philosophical premise was that type-identities are not necessary for a reduction. This debate is clearly unprincipled since one participant assumes type-identities for reducibility whereas the other participant simply dismisses this requirement. Without the context of a full model of a reduction the point is moot. One can give good reasons for both views and the debate cannot be settled. For a lucid discussion of various classificatory issues surrounding models of reduction and the issue of bridge-principles, see Sarkar, 1992.
 4. Sarkar (1998, Ch. 3) has developed this distinction in some detail in the context of genetics. The distinction goes back to Nagel's (1961, pp. 358-366) treatment of reduction.
 5. The situation is actually more complicated than this. Even the realization of this experimental science fiction scenario would not alone provide conclusive evidence for or against the opponent mechanism. According to De Valois' model, inhibition corresponds in some cells with green and in other cells with its opponent type red. The opponent code is ambiguous with respect to hues. This entails that we need to know more than the neural

- “response code” to test the model. For a systematic treatment of this point in the context of a critique of the explanatory concept of an “opponent code,” see Seppalainen, 2002.
6. For a direct assessment of the empirical evidence, see Seppalainen, 1999, Ch. 4-6.
 7. For a lucid explication of this “parity argument,” see Oyama, 2000b.
 8. Those who attempt to reserve a theoretical role for genetic coding tacitly admit this point. One popular option is to conceive the coding relation not as a mapping between DNA and traits but as a mapping between the order of amino-acids and the DNA triplets (e.g., Godfrey-Smith, 2000). Although the coding relation is only one among many relations leading up to a trait it is argued to be a special one because it is analogous to “reading.” In this interpretation, genetic codes escape the charge of a pseudo-explanation. But it is important to note that, in this interpretation, genetic coding becomes an explanatory notion because it is *not* the only explanatory factor for a trait.
 9. The *locus classicus* is Mill’s *System of Logic* (1874, pp. 533-538).
 10. The looks-like methodology is itself a consequence of this pragmatic approach to causation. The current methods require it since neuroscientists cannot directly intervene with the brain of a perceiving human on the neuronal level. In the color case, for example, the neurophysiological results are from our distant cousins, the macaques. The neuroscientists employ the looks-like methodology in order to correlate the data from two subject types, the macaque and the human, because of a lack of correlation data between the neuronal and the perceptual level in human subjects (for an elaboration, see Seppalainen, 2002).

REFERENCES

- Churchland P. M. (1985), “Reduction, qualia, and the direct introspection of brain states,” *Journal of Philosophy* **82**, pp. 9-28.
- Clark A. (1993), *Sensory qualities*. Oxford: Clarendon Press
- Clark A. (1996), “True theories, false colors”, *Philosophy of Science* **63** (Proceedings), pp. S143-S150.
- Dedrick D. (1996), “Can colour be reduced to anything?”, *Philosophy of Science* **63** (Proceedings), pp. S134-S142.
- De Valois R. I. (1965), “Behavioral and electrophysiological studies of primate vision”, in W. D. Neff (ed.) *Contributions to sensory physiology* Vol. 1. New York: Academic Press, pp. 137-178.
- De Valois R. I., Abramov I., and Jacobs G. H. (1966), “Analysis of response patterns of LGN cells”, *Journal of the Optical Society of America* **56**, pp.

966-977.

- Godfrey-Smith P. (2000), "On the theoretical role of 'genetic coding'", *Philosophy of Science* **67**, pp. 26-44.
- Grünbaum A. (1984), *The foundations of psychoanalysis: A philosophical critique*. Berkeley: University of California Press.
- Grünbaum A. (1993), *Validation in the clinical theory of psychoanalysis*. Madison, Connecticut: International Universities Press.
- Hardin C. L. (1993), *Color for philosophers: Unweaving the rainbow* (2nd revised edition). Indianapolis: Hackett
- Hartline, H. K. (1938), "The response of single optic nerve fibers of the vertebrate eye to the illumination of the retina", *American Journal of Physiology* **121**, pp. 400-415 (reprinted in F. Ratliff (ed.), *Studies on excitation and inhibition in the retina: A collection of papers from the laboratories of H. Keffer Hartline*. New York: The Rockefeller University Press, pp. 157-172).
- Hilbert D. (1987), *Color and color perception: A study in anthropocentric realism*. Stanford, CA: Center for the Study of Language and Information.
- Hurvich L. M. (1981), *Color Vision*. Sunderland, MA: Sinauer.
- Hurvich L. M. and Jameson D. (1957), "An opponent-process theory of color vision", *Psychological Review* **64**, pp. 384-404.
- Lennie P. and D'Zmura M. (1988), "Mechanisms of color vision" , *Critical Reviews in Neurobiology* **3**, pp. 333-400.
- Lewontin R. C. (2000), *The triple helix: Gene, organism, and environment*. Cambridge, MA: Harvard University Press.
- Mill J. S. (1874), *A system of logic* (8th edition). New York: Harper and Bros.
- Nagel E. (1961), *The structure of science*. New York: Hatrcourt, Brace & World.
- Oyama S. (2000a), *The ontogeny of information* (2nd enlarged edition). Durham, NC: Duke University Press
- Oyama S. (2000b), "Causal democracy and causal contributions in developmental systems theory", *Philosophy of Science* **67** (Proceedings), pp. S332-S347.
- Sarkar S. (1992), "Models of reduction and categories of reductionism", *Synthese* **91**, pp. 167-194.
- Sarkar S. (1998), *Genetics and reductionism*. Cambridge: Cambridge University Press.
- Seppalainen T. (1999), *The problematic nature of experiments in color science*. Unpublished Ph.D. dissertation, Department of History and Philosophy of Science, University of Pittsburgh.

- Seppalainen T. (2002), "Was there ever a physiological opponent color code?", in B. Saunders & J. Van Brakel (eds.), *Theories, technologies, instrumentalities of colour*. Lanham; Maryland: University Press of America, pp. 261-279.
- Teller D. T. (1980), "Locus questions in visual science", in C. S. Harris (ed.), *Visual coding and adaptability*. Hillsdale, New Jersey: Lawrence Erlbaum, pp. 151-176.
- Teller D. T. (1990), "The domain of visual science", in L. Spillman and J. S. Werner (eds.), *Visual perception: The neurophysiological foundations*. New York: Academic Press, pp. 11-21.
- Thompson E. (1995), *Colour vision: A study in cognitive science and the philosophy of perception*. London: Routledge.