Constancy, Categories and Bayes:
A new approach to representational theories of color constancy

ABSTRACT
Philosophers have long sought to explain perceptual constancy—the fact that objects appear to remain the same color, size and shape despite changes in the illumination condition, perspective and the relative distance—in terms of a mechanism that actively categorizes variable stimuli under the same pre-formed conceptual categories. Contemporary representationalists, on the other hand, explain perceptual constancy in terms of a modular mechanism that automatically discounts variation in the visual field to represent the stable properties of objects. In this paper I argue that while the former view is unmotivated by empirical evidence, the later fails to account for inter- and intra-personal variability, the influence of expectations on constancy, and the systematic and normal failures of color constancy. A Bayesian approach that builds on the representational tradition in psychology solves both problems.

Consider the phenomenon of perceptual constancy: that objects appear to remain the same color, size and shape despite changes in the illumination condition, perspective and the relative distance. Philosophers often try to explain constancy by asserting that perceivers actively categorize non-constant stimuli under constant conceptual categories. Contemporary representationalists about perception like Jackson, Tye, and Bryne & Hilbert offer an alternative explanation. According to them, color constancy demonstrates that our visual system represents constant properties of objects. In the case of color
perception, the visual system represents the surface spectral reflectance. Variations in perceived constancy are a matter of misrepresentation; specifically, the visual system misrepresents changes in the illumination conditions as changes in the surface spectral reflectance. Representationalists delineate between misrepresentation and successful representation by appealing to nonoptimal viewing conditions (Tye, 1995, p. 100), *ceteris paribus* conditions that fail to obtain (Tye, 2000, p. 140), or abnormal causal histories (Jackson, 1998, p. 96-100). Representationalists do not seek to explain color constancy *per se* but establish a theory of perception that explains color constancy in terms of what psychologists call the visual system ‘discounting the illuminant’ to recover the surface spectral reflectance of objects.

Neither the former ‘categorical’ explanation nor the representationalists’ appeal to the psychological tradition of discounting the illuminant is satisfactory, although the latter enjoys more empirical support than the former. Three problems arise: first, variability in color constancy, both inter- and intra-personally, is a phenomenal fact; second, while expectations do not determine color appearance, they really do influence perceived constancy; and third and perhaps most importantly, constancy fails in a systematic, predictable and quite ‘normal’ way. By relying on ‘cognitively impenetrable’ automatic, modular mechanisms, representational theorists underplay the very real influence of conscious processing on perception. Moreover, by merely labeling illusion as ‘misrepresentation’ and ‘abnormal’, they fail to provide a satisfying account of the *how* of normal, systematic illusion, content to merely hold *that* illusion occurs.

While the representationalist approach fails to adequately explain systematic and normal variation and failure, the categorical approach takes these variations as
fundamental, leaving successful constancy inadequately explained. By extending the psychological theory of ‘discounting the illuminant’ in terms of Bayesian models, I argue that we can solve both problems. The resultant view is broadly representationalist, but allows for a precise specification of how inter- and intra-personal variation may occur, how constancy failure can be both systematic and normal, and how expectations may influence perception. But it does so without making error the rule, as the categorical theorists do.

I am not concerned here with views that seek to deny perceptual constancy as a phenomenon, as the sense-data theorists did half a century ago. I am concerned with two types of explanations of color constancy. The representationalist approach posits that the visual system automatically discounts changes in the illumination conditions to produce stable representations of object color. The categorical approach posits that active categorical interpretation plays an integral role in the visual system’s ability to produce stable representations of object color.

This paper has three sections. In the first, I will state what I take to be the nature of the phenomenon in question, being careful to delineate it from tangentially related topics in contemporary philosophy of mind, and discuss some of the oft-abused historical record. In the second, I summarize the mechanism of the most promising categorical view found in the contemporary literature—Evan Thompson’s ‘enactive’ view—and consider two arguments in its favor: first, the argument from task- or judgment-dependence; and second, the argument from infants and animals. In the final section, I turn my attention to representationalism. I argue that while representationalism is on the right track, at least insofar as it builds on the well established psychological theory of
‘discounting the illuminant’, the simple version adopted by Jackson, Tye, and Byrne & Hilbert cannot explain the systematic, predictable and quite normal failures of color constancy. A more sophisticated Bayesian model can.

One of my secondary goals in writing this paper is to correct some of the mistakes I see in the understanding of color constancy research perpetuated in the philosophical literature. The best empirical research on the matter is, as I hope to show, neither categorical nor strictly representational. It is broadly representational, but is also consistent with many of the insights of the categorical theorists like Evan Thompson. I will limit my explication of the extraordinarily complicated 140-year history of color constancy research, but I hope the reader will forgive the occasional explication of empirical work in support of this secondary goal.

1. Preliminaries and History

Allow me to illustrate the phenomenon of color constancy with an example: bathed in the fluorescent light of my basement office, I glance at the cover of a recent issue of *Philosophical Psychology*. It looks green. Not the green of maple leaves in July or of the lawns at the All England Tennis Club, but green nonetheless. Needing some fresh air, I grab the issue and head for the quad. There, *Philosophical Psychology* looks, well, green. Again, it’s neither the green of the oaks nor of the grass. But it is still that same odd shade of green. At dusk, it still looks green. At dawn the next morning, still green. Why? The lighting conditions vary dramatically between those four occasions. The light reflecting off the cover of *Philosophical Psychology* varies in just the same manner. Why, then, does the cover of this journal remain the same distinctive green in each case?
Two explanations strike us as plausible. First, *Philosophical Psychology* appears to be a stable green because colors are the stable properties of objects—my visual system automatically discounts variations in the illumination in order to recover stable, mind-independent properties. This is an instance of the representationalist explanation. Second, *Philosophical Psychology* appears to be a stable green because I continually categorize it under the same, pre-formed perceptual concept of green independent of the physical facts of the matter—I impose stable, pre-existing color categories upon unstable, variant mind-independent properties. This is an instance of the categorical explanation.

By ‘color constancy’ I refer to the ability of our visual system to stabilize the perceived colors of objects independent of changes in the illumination conditions under which those objects are viewed. All colored objects have stable dispositions to reflect a percentage of light according to the light’s wavelength. We call these dispositions the object’s ‘spectral reflectance’. Transparent, or uncolored, objects have no such spectral reflectances. The spectral composition of the light reflected by an object (hereafter the ‘color signal’) varies as a function of that object’s spectral reflectance and the spectral composition of light incident to the object. The color signal when multiplied by the sensitivity curves of the three types of cones in a normal (human trichromatic) retina is called the ‘tristimulus values’. Two stimuli with the same tristimulus values in a given lighting condition will appear to be the same color in that lighting condition (although they make look different in different lighting conditions). It is important not to confuse these two. The color signal is a property of the light; the tristimulus values represent that light once it is processed by a normal human perceiver. It is possible for two objects to
have different color signals and the same tristimulus values, if the difference in the color
signals is not detectable by the sensitivity curves of the photoreceptors in a normal retina.

As I look at the cover of *Philosophical Psychology* in my office, the color signal
is the product of the journal cover’s spectral reflectance and the spectral composition of
fluorescent light. The tristimulus value is that signal processed by my (theoretically
abstracted) retina. In the quad, the color signal is the product of the cover’s spectral
reflectance and the spectral composition of direct sunlight. The journal’s spectral
reflectance and my retinal are largely stable between my office and the quad, so most of
the differences between the resultant color signals result from the differences between the
spectral compositions of fluorescent light and sunlight.

Color constancy is a robust phenomenon, but it is not perfect. The cover of
*Philosophical Psychology* does not look that same, unique shade of green inside a dance
club, nor does it retain its characteristic hue when viewed under the sickly-yellow light of
a street lamp. When the book is held at just the right angle to the incident light, it takes on
a gradient of both lightness and hue. If the perceived color of an object varies along with
the illumination conditions, we have a failure of color constancy. If there is no such
variation (or little enough variation), color constancy is successful. As a first pass, we
can say that ‘good’ or ‘successful’ constancy is shown if the variation between the
perceived color in two lighting conditions is significantly less than the variation between
the lighting conditions themselves.

1.1. **What is at stake in this debate?**

Consider size constancy: that objects appear to remain the same size even when
viewed from a distance. Locke proposed that constancy was explained by “habitual
custom,” while Bishop Berkeley found that “the estimate we make of the distance of objects considerably remote is rather an act of judgment grounded on experience than of sense.” (1963, §3, p. 20) Indeed, Berkeley’s categorical explanation provides the crucial premise in most of his arguments for mind-dependence of reality.\(^6\)

Few philosophers today would endorse Berkeley’s argument for the mind-dependence of size. But there are many who wish to hold onto the mind-dependence of color. It follows, therefore, that these color anti-realists must seek a perceptual property that is shared by colors but not by mind-independent ‘primary’ properties like shape and size upon which they can ground the distinction.\(^7\) Color constancy presents the opposite problem for the anti-realists. Constancy is a perceptual property of primary qualities. Shapes remain perceptually constant as we move around them. People do not seem to shrink as they walk away, and coins laid out on the desk look like round coins seen from an oblique angle, not ovals seen in perpendicular portrait. If colors are constant, like shapes and sizes, then the simplest explanation appears to be that colors are primary qualities—that is, if the representational explanation, and not the categorical explanation, is correct.

This is not to say that the representational view articulated here determines the primary quality view with absolute certainty. Indeed, the primary quality theorists cited above claim that the primary quality view is the simplest explanation given the representational approach to constancy. They do not claim that it is logically entailed by constancy. It is logically possible to be committed to the kind of Bayesian representationalism I advocate and reject the primary quality view of colors.\(^8\) But it is, I believe, an overly complicated position.
1.2. A Little History

Nearly a century after Berkeley’s initial judgment-based categorical explanation of size constancy, Helmholtz theorized that colors remain constant in changing light conditions because we judge them to be so; however, unlike Berkeley, he theorized that the relevant judgments occur instantaneously and are inaccessible to conscious reflection (see, e.g. 1924, p. 286). Helmholtz later abandoned his doctrine of ‘unconscious inference’ as a viable explanation for the phenomena holding first that unconscious reasoning ‘like inference’ takes place and, ultimately, that color constancy is determined by the relationship between the surface reflectance and the illumination. Indeed, he went so far as to hypothesize that the average reflectance seen in the visual field is identified as white, and all other colors are calculated in relation to that. Failures of color constancy—particularly color contrast and colored shadows—are explained in terms of mis-identifications of or changes in that which is identified as white.

The next great theorist of color constancy was David Katz, who theorized that color constancy was a matter of psychological recognition of ‘abnormal’ viewing conditions which in turn lead to physiological adjustment (1911). In 1929, Gelb conducted a famous experiment to test this theory, in which the participant was seated in a darkened room across from two disks, a shadowed white disk and an illuminated black one. The participant viewed each disk through a ‘reduction screen’—a large screen with a hole small enough that the viewer cannot see through the hole, but only see the hole. The participant then adjusted the light illuminating the black disk until it appeared approximately the same as the white. When the screen was removed, the black disk immediately changed to dark black, and the white appears unchanged. According to
Gelb, the black, illuminated disk was in a ‘normal’ illumination condition, while the white, shadowed disk was ‘abnormal’. Hence, color constancy cannot be a matter of psychological adjustment towards ‘normality’ (Gelb, 1938, p. 204-5).

This experiment led Katz to abandon his categorical view in favor of what is now called the ‘articulation hypothesis’. According to Katz, color constancy varies with the degree to which the object bearing that color is articulated in the visual field as an object. The objectivity of an object is, for Katz, a phenomenal fact. This new theory is no longer categorical, as the visual system automatically recovers the colors of objects based on quantifiable properties of the visual field.

1.3. Tangential Distinctions.

A number of distinctions currently debated in contemporary philosophy of mind might be easily confused with the distinction between the representational and the categorical explanations of perceptual constancy. Determining a solution to the current problem may well influence one or more of these research programs, but I see no necessary connection between any of these and the distinction at hand.

First is the distinction between language-dependent and experience-dependent concepts. There are a few who would argue our perceptual systems are so influenced by the language that we literally cannot see color without having the relevant linguistic categories. Fortunately, these people are rare in this post-Berlin & Kay world. But they may still exist. This argument is not an issue here, as Evan Thompson, perhaps the leading categorical theorist, is also the leading theorist with respect to the color vision of non-linguistic creatures. The categorical view holds that color perception requires color concepts; not that color concepts require language.
Second is the distinction between conceptual and non-conceptual content. While this distinction has some contact with the issue at hand, the way in which the current debate about the distinction is framed—i.e. in terms of the ‘fineness of grain’ of experience—is inapplicable. It is an empirical matter just how finely color constancy cuts. When I see ‘that shade’ of green outdoors, and then ‘that shade’ of green indoors, any variation in the precise match is so small as to not raise to the level of conceptual processing for naïve observers, but that does not mean that there is no variation. Colors remain \textit{approximately} constant. They do not remain precisely constant. While it may be the case that the variation found between illumination conditions cuts finer than even the color vocabulary used in common parlance can account for, it does not cut finer than precise, scientific or artistic color vocabulary (e.g. the Munsell system).

Third is the distinction between color concepts as innate or learned. If ‘innate’ means biologically determined syntactic structure of perceptual color categories, then it could be related to the distinction I am interested in here. However, more often than not, this distinction is a red herring; for even if the categories were learned, they would be learned in response to regularities found in the natural world. And if they were biologically determined, they would be so wired by natural selection pressure imposed by regularities in the natural world. But in both cases, they need not be so. Just to prove the lack of necessary connection here: Hatfield (1992; 2003b) relies on the categorical view to discuss the biological function of color vision, while Hilbert (1992) and Kraft, Maloney & Brainard (2002)\textsuperscript{14} assert the biological function of color vision to motivate the representationalist view.
Finally, in the psychological literature, ‘categorical’ perception is often opposed to ‘continuous’ perception. Linguistic audition, for example, is categorical because we hear words in a sentence as discrete units of speech, rather than continuous flows of sound. There is little doubt that linguistic audition results from a mechanism of which processes such as expectations, learning and judgment are in an integral part—but the distinction between categorical audition of language and continuous audition of, say, white noise, is a distinction between the phenomena produced by such a mechanism, not between the mechanisms themselves. While there is probably a connection between the phenomena and the mechanism that produces it—i.e. categorical perception might require a categorical mechanism—it is by no means necessary. Moreover, color constancy is simply not a categorical phenomenon in this sense. Hues are not seen as discrete units, and their variations in lighting conditions do not jump over large qualitative gaps. My concern in this paper is a categorical mechanism of a continuous phenomenon, not whether or not colors are themselves categorical. I do find it helpful, however, to consider linguistic audition as a paradigmatic categorical mechanism in order to understand the lines of reasoning used by categorical theorists.

The distinction between categorical and representational theories of color constancy is perhaps best stated as ‘expectation-dependence’, as opposed to ‘mind-dependence’ or ‘judgment-dependence’. The question I am interested in is simply stated thus: does color constancy essentially depend on our expectation that colors in the world are stable, or is the stability of color presented to us as a fact of our experience? The categorical explanation holds that an active consciousness, at the level of agency, is a necessary part of the mechanism of color perception. The representational approach holds
that color constancy results from an automatic or modular mechanism, and no act of agency is required to perceive colors. The Bayesian model I advocate threads the difference between these two, advocating a mechanism that represents spectral reflectances by mathematical transformation on the tristimulus values, which is still open to the influence of expectations. Hence, Bayesian representationalism need not deny the role of conscious experience and expectations in the construction of unambiguous percepts from ambiguous stimuli as the words ‘automatic’ and ‘modular’ imply. But it does deny that these acts of will and/or unconscious expectations are an essential or integral part of the mechanism of perception itself.

2. The Categorical Explanation

Although I do not have the space here to discuss in depth the history of color constancy research in psychology, it offers us little reason to endorse the categorical view. For example, Hering explained constancy in terms of ‘memory color’:\(^\text{15}\) that the visual system stabilizes color appearance from memory and projects that color onto known objects. In 1911, Katz showed that color constancy holds for objects that have never been seen before, as well as random shapes that might have previously been seen bearing any color. This result was replicated by Fisher, Hull & Holtz in 1956 and Land in 1977.

Not only is memory woefully inadequate in explaining even the simplest color constancy phenomena, but also its only explanation of constancy failures must be in terms of misapplied or misremembered memories. These mis-memories will be, by the nature of memory, individually unique. And hence, the model simply can’t explain the systematic failure of color constancy.
The philosophical literature offers us little more insight. Alva Noë, for example, merely states: “The phenomenon of color constancy demonstrates the perceivers possess this sort of implicit knowledge.” (2004, p. 127) By ‘this sort’ of implicit knowledge, he means knowledge about the variability of color appearance in different lighting conditions learned, “unconsciously in childhood” (ibid). This appears to be a simple resurrecting of Helmholtz’s principle of ‘unconscious inference’, variations of which have been considered and rejected by Helmholtz himself, Katz in 1911 and 1935 and Land in 1977.\(^\text{16}\)

Let us consider the issue by returning to the most obvious case of categorical perception: linguistic audition. While a categorical perception like linguistic audition may or may not require a categorical mechanism like the kind we are considering here, a hypothetical categorical mechanism will almost certainly produce categorical percepts. As such, we can extract few lines of reasoning to support the categorical mechanism of color constancy from the analogy. First, linguistic audition is subject to cross-modal influence (i.e. listening to the mumbled lyrics of Joe Strummer of the Clash vs. reading the lyrics as they are sung). Second, linguistic audition generally depends on the task being performed (i.e. listening to the television vs. listening to one’s spouse). Third, auditory experiences should be measurably different for non-linguistic creatures such as infants and animals. By analogy, if color constancy is the result of a categorical mechanism, color constancy should be subject to cross-modal influence, dependent on the task being performed and/or measurably different for infants and animals. I know of no empirical work—or philosophic arguments—that approach the first line of reasoning. The later are worth considering.
The argument from task or judgment dependence, is embodied in Evan Thompson’s seminal book *Colour Vision: A Study in Cognitive Science and the Philosophy of Perception*:

[Arend & Reeves’] results can be taken to indicate, first, that observers can perceive both surface chromaticity and illumination conditions, and second, that colour constancy is not automatic or modular (in, say, Pylyshyn’s (1984) sense of being ‘cognitively impenetrable’), but rather strongly task- and judgment-dependent. (1995, p. 197)

The argument from infants and animals does not appear explicitly in the work of either Thompson or Noë, but it is prevalent enough to require consideration. Unfortunately, the empirical evidence for both of these is weak, and the arguments uncertain. Before tackling their arguments directly, it is worth taking a moment to clarify exactly what mechanism of color perception the categorical explanation offers.

2.1. The Categorical Mechanism: Thompson’s enactive view

While philosophers have often expounded a categorical explanation of color constancy with little, if any, empirical foundation, Evan Thompson has presented an empirically informed theory that instantiates the categorical explanation; and it is worth the time to consider it carefully. First and foremost, Thompson’s mechanism is naturalistic. Second, it is committed to the view that “the animal is not simply a collection of neuronal and psychophysical processes, but most properly also a perceiving and acting subject.” (216, italics in original) And third, it is committed to the thesis that the animal and its environment is best described as “jointly composing a larger system—an animal-environment ecosystem.” (218)

14
Colors must be, according to Thompson, “considered as an ecological-level dispositional property. At the ecological level, colours... are tied to affordances and effectivities: they indicate what the environment affords for the animal and what the animal can effect in its environment.” (224) They can take three forms: a property of surfaces, ambient light in the aerial or aquatic media, or a property of ecologically significant light sources such as the sun or bioluminescent organs (224). In each of these later three forms, color corresponds to a spectral reflectance function; hence these latter three forms fail to distinguish Thompson’s view from representationalism. The only significant point of distinction is the one quoted above. According to Thompson, colors must be understood as affording the organism abilities in the environment. In the case of color constancy, the ability afforded to the organism is clear: color constancy allows the organism to recognize the same object in different lighting conditions. This recognition, however, is a matter of categorical perception as it “serves to integrate a physically heterogeneous collection of distal properties into a small set of visually salient equivalence classes.” And the ‘salient equivalence classes’ are, for humans, the four opponent color classes (called ‘super-determinables’ by Bryne & Hilbert in their 1997) reddishness, yellowishness, greenishness and bluishness.

Thompson’s mechanism is now complete. Colors are those properties that afford us ecological-level categorization under four super-determinable color categories. These color categories, or their equivalents, are not linguistically dependent as they can be possessed by creatures as simple as honeybees (although the categories will be different). The mechanism of color representation, then, is the computational mechanisms of
contemporary psychophysics with the addition of an agent who actively categorizes distal stimuli under pre-existing color categories.

Active categorization at this point in the mechanism is problematic in two respects. First, color constancy fails, but does so in a regular, systematic way. While there are inter- and intra-personal differences in color constancy, there are certain well-known conditions, such as the Gelb effect, in which all perceivers suffer the same constancy failure. A good theory of color constancy should be able to predict not just \textit{when} constancy will fail, but precisely \textit{how}. In the case of the Gelb effect, a good theory should be able to predict precisely what shade the disk should appear when the viewing condition changes. Thompson’s mechanism assigns this to active categorization. Subjective agents, who play the crucial role in perception for Thompson are, by definition, unpredictable. Hence, one cannot make a prediction of color appearance when misrepresentation occurs. Second, Thompson has not given us veridicality conditions other than whatever affords the organism consistent recategorization. Therefore, we cannot explain why an agent chooses to categorize a given object as falling under a certain hue category \textit{other than} that that agent has done so previously. This leads to a digression.

2.2. \textbf{Line of Reasoning 1: task or judgment dependence}

The wall directly in front of me does not appear to be uniformly beige: the lamp on my desk casts a slight greenish hue in a semi-circle behind it. The florescent lights in the ceiling create a vertical gradient transitioning from very bright at the ceiling to rather dark at the point where the wall meets the floor. And finally, the whole wall is subject to the minor variations in apparent hue where the sunlight streaming through the window in
the center of it reflects off the various historical works of experimental psychology propped up on my desk.

Painters are trained to notice these differences by adopting what Gelb called ‘a critical viewpoint’. The question for us is whether the ability to notice such variations reveal a categorical mechanism of color constancy or merely difference of attention. The differences in apparent hue that artists are adept at noticing are minor in comparison to the differences in illumination studied by those interested in constancy phenomena. In 1977, Edwin Land showed that a red patch and a green patch can, in different lighting conditions, reflect the same composition of light, yet still appear red and green. No artist has ever claimed to be able to see red when looking at a green patch merely by adopting a critical viewpoint. On the other hand, no one can deny that painters masterfully replicate the variations in perceived color that result from variations in illumination. Indeed, it is precisely this ability that allows for the rich representation of colored objects on canvas.

The question for the categorical view is whether this task- or judgment-dependent ability, which can be perfected with training, motivates the thesis that active categorization under pre-existing hue categories is necessary for color constancy.

Thompson indirectly argues for a categorical mechanism by asserting that the experimental work of Arend & Reeves conducted in 1986 points to the task- or judgment-dependence of color constancy; and hence, constancy is not ‘automatic’ or ‘modular’ as his opponents the simple representationalists would have.¹⁸

Let us proceed carefully. Arend & Reeves are not out to establish the categorical explanation. Their paper is part of a long debate between those who, like Helson & Judd, think that color constancy results from a low-level simultaneous adaptive mechanism that
is akin to color contrast and those who think that color constancy is a distinct phenomenon that must be explained in terms of computational processes. Thus, their experiments are designed to show whether color constancy results from an ‘adaptive’ or ‘simultaneous mechanism’. While Arend has continued this research since 1986 with the publication of at least two more papers on similar experiments (Arend & Goldstein, 1987; Arend, Reeves, Schrillo, & Goldstein, 1991), the paper most commonly discussed—and the one cited by Thompson—is his first. And hence, that is the one to which we’ll turn our attention. The conclusion of Arend & Reeves 1986 is that:

- mechanisms such as simultaneous color contrast, which inflexibly discard, through a normalization process, information about abrupt spatial illumination changes, make little contribution to color constancy within a single scene. Humans can, however, compute approximate chromatic reflectance information in such scenes while attaining some illumination information (1986, 1749).

Therefore, color constancy is an adaptive mechanism. It is the data behind this last sentence that gives hope to the categorical explanation. Read without theoretical commitment, it appears to endorse something like the ‘critical standpoint’ point made by Gelb. But given that this is the only evidence offered in favor of the categorical explanation, we must look at the experiment in some detail.

Three participants (the two authors and one naïve observer) were adapted to an ambient light for three minutes, and then asked to match a test patch on the left to one of two continuously displayed standard patches on the right. Both single color and Mondrian displays were used. The participants were given one of two tasks. In the first,
hereafter the ‘hue-match task’, participants were instructed to “match the hue and saturation of the test patch to that of the standard patch.” In the second, hereafter the ‘paper-match task’, to make it look “as if the two patches were cut from the same piece of paper.” (p. 1747)¹⁹

According to Arend & Reeves’ results, participants were much less constant when performing the hue-match task than the paper-match task. In the paper match condition “two of the three subjects had sufficient knowledge of the color that the paper should take on in the test illuminant to produce approximately color-constant settings.” (p. 1747) The authors do not mention if both of these subjects were themselves, or if only one of them had this ability.

From this, Arend & Reeves reason that color constancy results from two distinct mechanisms. It might occur because the visual system contains a mechanism that:

- adjusts the response [of the photoreceptors] to cancel the proximate stimulus differences produced by the illuminant change. For example, a paper that looks unique yellow under direct sunlight might continue to look unique yellow in the greenish reflected illuminant under a tree (1749, italics mine).

Or color constancy might occur because colors change, but are thought not to: “The paper that looks unique yellow under direct sunlight might look greenish yellow under the tree and yet might be clearly identifiable as a yellow paper” (1749, italics mine). According to Arend & Reeves, the failure of color constancy in the hue-match condition demonstrates that constancy is not a matter of “a sensory mechanism that adjusts the response to cancel the proximate stimulus differences.” Therefore, by
disjunctive syllogism, color constancy is a matter of identifying objects as being such-and-such a surface. Hence, color constancy is successful in the paper-match condition because it is a matter of seeing the paper “as an object of the same surface color.” (1749)

It is difficult, given the bipolarity of their dilemma, to get a handle on exactly what is meant by being ‘clearly identifiable as a yellow paper’ without it ‘looking yellow’. And given that the mechanism they reject is not explicated in terms used historically such as ‘physiological’ or ‘psychological’, ‘categorical’ or ‘phenomenal’, it is difficult to say exactly what Arend & Reeves have shown. One could argue that Arend & Reeves’s argument is consistent with Katz's rejection of his own theory of 1911: that color constancy is a matter of seeing colors as properties of surfaces, not of a psychological adjustment towards ‘normality’. After all, for them, constancy is a matter of perceiving surface color. On the other hand, one could argue that Arend & Reeves have shown that perceivers can be induced to make different judgments about color in contrived experimental conditions. Whether those judgments override the phenomenal experience or determine it has not been investigated in this experiment—and that is what is at issue for this paper.  

It appears that we have a dilemma for the categorical view. Either Arend & Reeves have demonstrated a minor point that individuals, with training, can attend to minor shifts in color appearance due to changes in the ambient light; or their results are merely tangential to the issue at hand. Either way, it certainly fails to motivate the categorical explanation. Moreover, as Thompson’s argument is a disjunctive syllogism on the notions of ‘automatic’ or ‘modular’, even if their experiment was externally valid, it would have no impact on the sophisticated Bayesian representationalism I advocate.
2.3. Line or Reasoning 2: Infants and Animals

In his classic *Gestalt Psychology*, David Katz hypothesized that “the tendency to form objects would become operative in the consciousness of a child from the start, even without experience of any sort.” (p. 22) Here, Katz is providing us with a crucial experiment to test the phenomenal explanation of object perception—and it can be extrapolated to color constancy.\(^{21}\) Very young infants, it could be argued, lack the categorical concepts required to make colors constant. If young infants are not color constant, it might be the case that constancy is categorical. On the other hand, the same might be said for animals; and the empirical evidence that goldfish are color constant has not been called into doubt by those advocating the categorical view. In fact, animal constancy appears to have been known since the time of the Gestalts; Gelb reports that Katz himself showed that chickens and apes were color constant as early as 1921, and goldfish in 1923.

Animal color constancy began to be re-established starting in 1980, when Neumeyer demonstrated color constancy in the honeybee. Werner, Menzel & Werhahn confirmed this result in 1988. Dörr & Neumeyer demonstrated color constancy in goldfish in 1996 and established its limits in 2000. Tritsch demonstrated it in cats in 1993. To hold onto the categorical explanation of color constancy, one would have to hold onto the thesis that goldfish have innate color concepts under which they conceptually categorize the objects in their environment. That seems a bit of a stretch. One could simply deny the empirical evidence that animals are color constant, but that is a strategy that Thompson, at least, is unwilling to pursue. So much for animals. But what of infants?
In 1989, Dannemiller\textsuperscript{22} conducted a study comparing the color constancy of 9- and 20-week old human infants (1989b). Both groups were shown computer-simulated pairs of natural surfaces with complementary reflectance functions illuminated by two different phases of sunlight. Dannemiller concludes that:

The 20-week-olds but not the 9-week-olds showed evidence of color constancy. The 20-week-olds responded with increased attention only to simulated changes of surface reflectance and not to simulated changes of the illuminant alone. However, 9-week-olds responded with increased attention to simulated changes either of the illuminant or of the surface reflectance (p. 180-1).

Here, we have an apparently significant result, although I have yet to find a citation of this study in the philosophical literature. There are, perhaps, a couple of concerns. First, infants’ cones are not fully developed until at least 4 months of age, so the difference in the results may be due to differences in cone development.\textsuperscript{23} Second, infants’ memory abilities develop rapidly during this time. And we know that adults regularly select more saturated samples after a 10 second delay than the original test patch.\textsuperscript{24} Given that the test stimuli were presented more than ten seconds after the target stimulus, the variation in results may be explainable in terms of a difference in the memory abilities of nine and twenty week-old infants. These critiques were considered and rejected by Dannemiller as not sufficient to explain the failure. I am less certain.

Research on color constancy in infants continues, but these challenges to the experimental design make it difficult to tease apart the mechanisms in question. Given
the lack of citations to this research in the philosophical literature, I will leave the topic here.

3. **Representationalism**

Representational theories of perception provide an alternative model of color constancy, according to which the visual system recovers, or represents, the spectral reflectance of objects from the color signal. Representational theories of perception have been the subject of much discussion in the philosophical literature, so I will not rehash their vices and virtues here, except to point out two problems with the way they are usually presented: first, by relying on ‘cognitively impenetrable’ automatic, modular mechanisms, representational theorists underplay the very real influence of conscious processing on perception; and second, they tend not to explain the how of illusion, merely labeling it ‘misrepresentation’ and blaming it on nonoptimal viewing conditions, non-obtaining *ceteris paribus* conditions or abnormal causal histories. These faults can be rectified without resorting to ill-defined notions like ‘agency’, I suggest, by adopting a Bayesian model of color constancy, according to which misrepresentation is the visual system’s best guess given invalid information.

3.1. **Recent Approaches to Color Constancy: Discounting the illuminate**

Over the past twenty years, research into color constancy has flourished. This resurgence was sparked by the publication of Maloney & Wandell’s “Color constancy: a method for recovering surface spectral reflectance” in 1986.\(^{25}\) The authors theorize, to summarize, that the visual system recovers the spectral reflectance from the color signal. The color signal, you will recall, perfectly confounds the spectral reflectance of the object
with its illuminant. Thus, the visual system must discount the illuminant from the color signal. The visual system recovers, or as representationalists prefer ‘represents’, the spectral reflectance successfully insofar as it successfully recovers or represents the composition of the illuminant to be discounted. And it follows that constancy fails when changes in the illumination are represented as changes in the spectral reflectance. A good model of color constancy should tell us not only in what conditions color constancy will succeed and fail, but also precisely how the variations in the illuminant will be represented as properties of objects in the conditions likely to produce constancy failure.\textsuperscript{26}

While there is a century of experiments and models based on recovering the illuminant,\textsuperscript{27} the most promising demonstration of the theory’s predictive power comes from the ‘nearly natural viewing condition’ experiments conducted by David Brainard and colleagues. Computationally, the model, hereafter the ‘equivalent illuminant model’, tested in these experiments is calculated by comparing the tristimulus values of the test color in a known ‘original’ illuminant (C\textsubscript{1}), with the tristimulus values of the match color (C\textsubscript{d}) and the test color if it were illuminated by the illuminant illuminating the match color (C\textsubscript{2}). If C\textsubscript{d} and C\textsubscript{2} are identical, we have success. If they are different, we have failure. Thus, constancy (C\textsubscript{I}) is given by the absolute value of the difference between the C\textsubscript{1} and C\textsubscript{d} divided by the absolute value of the difference between C\textsubscript{1} and C\textsubscript{2}. If C\textsubscript{d} and C\textsubscript{1} are identical (the match and the test patch under the \textit{original} illuminant), the numerator is 0 and hence the value of C\textsubscript{I} is 0 regardless of the value of the original illuminant. If C\textsubscript{d} and C\textsubscript{2} are identical (the match and the test patch under the \textit{experimental} illuminant), they cancel each other out of the equation, leaving C\textsubscript{1} to be
divided by C1, and hence CI equal to 1.\textsuperscript{28} Successful constancy is indicated by a value greater than 0.8.\textsuperscript{29} This model does not seek to explain how the visual system obtains an estimate of the illuminant—\textsuperscript{30} that is a more complicated question. The model merely seeks to establish that color constancy is a matter of discounting the illuminant.\textsuperscript{31}

One could argue, for example, that the visual system obtains this estimate via a categorical mechanism—specifically, an explicit judgment is made regarding the illuminant, against which spectral reflectances are then calculated. Unfortunately, there is little evidence that this is the case. Rutherford & Brainard asked seven naïve observers to match the illuminant in a target chamber to the illuminant in a test chamber, and then to adjust the lightness of a test patch in the test chamber to match a similarly positioned target patch in the target chamber. If the illumination estimates underlying color constancy are explicit judgments, these matches should be perfect. They were not. It follows that whatever the mechanism of illuminant estimation, it is not accessible to judgment (2002). Research on color constancy becomes research on how the visual system estimates the illuminant. And, as it turns out, that is not nearly as simple as it may first appear, as the visual system does not have direct information about the illuminant.

Models have been proposed to solve this problem in one of three basic ways. First, what we might call the ‘classical models’ hold that the visual system estimates the illuminant by adaptation to local conditions.\textsuperscript{32} Second, ‘spatial mean’ models hold that the visual system takes the average of reflectances as the basis for its estimate of the illuminant. Helson & Judd’s theory of color conversion and Land’s ‘retnix’ theory can be seen in this tradition. Also, researchers such as Gilchrist (1980) and Brainard & Wandell (1986) who have emphasized scene complexity as a contributing factor for color
constancy generally do so because complexity of scenes yield better averages. And third, ‘anchoring’ models hold that the visual system uses the part of the visual field as a heuristic to estimate the illuminant (Land & McCann, 1971, Cataliotti & Gilchrist, 1995 and Gilchrist et al., 1999).

All the models propose mathematical transformations from photosensor responses to spectral reflectance profiles that approximate the spectral composition of the illuminant, which in turn are used to discount the illuminant from the color signal to recover the spectral reflectance. In short, they are broadly representationalist. And all of these work well in certain scenarios and poorly in others. The Bayesian approach is slightly different. It places the visual system in the role of an ‘ideal perceiver’, analogous to an ‘ideal decision maker’ in decision theory, which opts for the most likely representation of the illuminant given all the information available. On this model, constancy is successful when the cues used by the visual system to estimate the illuminant give reliable information about the illuminant (i.e. ‘valid’ cues), and it fails when the cues are not reliably correlated with the illuminant (i.e. ‘invalid’ cues).

Thus, the visual system may opt for the most intense image region as the best estimate of the illuminant in certain ‘unnatural’ laboratory conditions such as those created by Land & McCann, and it may opt for the spatial mean as the best estimate of the illuminant in ‘nearly natural viewing conditions’ such as those created in Brainard’s lab. It may represent a spectral reflectance as a source of illumination if no best estimate is available—i.e. when viewing objects through Katz’s reduction screen, as in the Gelb effect. But in each and every case, it chooses the property or properties of the visual scene most likely to correlate with the illuminant and forms an estimate of that illuminant.
from the information available about that property or properties. In an important forthcoming paper, Brainard et al. quantify the Bayesian model of the visual system and show evidence of its predictive power with respect to both constancy success and failure (Brainard et al., Forthcoming). 34

3.2. Representational Bayesianism

All representational theories of perception must analyze how successful, veridical representation is distinguished from unsuccessful, ersatz representation. The problem is a difficult one. Consider some of the constraints on an adequate theory. First, successful states cannot be distinguished from unsuccessful states on the basis of phenomenology. When suffering an illusion with respect to color appearance, one cannot be aware of that fact through mere introspection. Second, there is general agreement about certain misrepresentations: everyone sees afterimages, is subject to the McCollough effect and the Gelb effect. Third, there is nothing ‘abnormal’ about the perceivers or the viewing conditions of the Gelb effect—in fact, it is the normal viewing condition that appears to be the case of misrepresentation. Fourth, as in the Gelb effect, misrepresentation may occur because of the presence of objects in the visual scene that do not appear, prima facie, to interfere with the physical facts of the matter. If all other things need to be equal for representation, all other things may need to be equal. Finally, constancy failures do not obviously involve some causal interference or abnormal causal history in the way that simultaneous or successive contrast effects do.

According to Bayesian models of perception, the visual system calculates the likelihood of representation—as all Bayesian models calculate—by multiplying the probability of that representation given the stimulus by the prior probability of that
representation, and then dividing that by the probability of the stimulus itself. In the context of discounting the illuminant, the representation is a spectral reflectance function that represents the illuminant, and the stimulus is the retinal activity. The visual system opts for the most likely illuminant given retinal activity, constrained by the prior probability of that illuminant actually occurring. As in the case of linguistic audition, the stimulus constrains the set of possible interpretations that can be imposed on it by a perceiver. Rarely, if ever, can the stimulus determine its interpretation.

In the first step, the model does not calculate the probability of a given spectral reflectance on the surface of objects—it calculates the probability of a given illuminant that is then used to calculate the spectral reflectances of objects in the visual field. In the second step, the Bayesian law appears again, calculating the probability of a spectral reflectance given the illuminant and color signal constrained by the prior probability of that spectral reflectance. The Bayesian approach thus allows for calibration of the model to ‘real-world’ illuminants and ‘real-world’ reflectances—specifically, the kind of illuminants and reflectances to which we would expect a naturally evolved visual system to be tuned. And it proposes a concrete, testable method that may account for inter- and intra-personal variation in color constancy: inter- and intra-personal variation in the priors. This approach has three significant advantages over both the categorical explanation and simple representationalism.

First, the spectral reflectance calculated by the visual system when information about the illuminant is unreliable will deviate from the actual spectral reflectance in exactly the same way that the illuminant estimate deviates from the actual illuminant. Thus, if the visual system incorrectly assumes that the black disk illuminated by a
spotlight in a darkened room is actually illuminated by the same darkened ambient light, it will misrepresent the disk as being much brighter than the actual. When the visual system gets reliable information about the spot (i.e. when a white disk is placed in the light), it will correct itself automatically by opting for the more likely representation. The switch will occur when the information available makes the current interpretation significantly less probable than the alternative interpretation. Hence, one would expect mutually incompatible representations, from which one cannot access the other by sheer act of will. And that is consistent with the phenomenology of the experience.

Second, while the best fitting model to date assumes priors that represent the statistical structures of illuminants in natural images, one could easily extend the model to allow cognitive influence—expectations or prior experience—to affect the priors. Not only could there be variation in the priors used between subjects, but they could vary within a subject over time. Such variations would not be enough to override huge gaps in the likelihood of a particular representation (say, seeing red as green by an act of will), but it could easily explain the kind of minor influences we see in painters, or was demonstrated by Dunker in 1939.35 I’m not arguing here that anyone has attempted such explanation in the psychological literature. I am arguing that this approach provides a clear and concrete mechanism that can explain such variation as opposed to the unclear and diaphanous alternative explanation from ‘agency’ or ‘active perception’. At the same time, the model is consistent with Thompson’s intuition that hue classifications may be influenced by prior ‘super-determinable’ categories. But again, the Bayesian approach models these categories in a precise, quantifiable way without equivocatory notions like ‘concept’.

29
Third, Bayesian models explain failure—misrepresentation—through *normal* causal histories. Failure is therefore generalizable enough to provide the basis for agreed upon, but false, cases of misrepresentation. This is a significant advantage over simple representationalism that relies on nonoptimal conditions, non-obtaining ceteris paribus conditions, or abnormal causal histories to explain misrepresentation. After all, there is nothing ‘abnormal’ about constancy failure. It happens in the natural world quite frequently, and there may be good biological reason for it. All other things being equal, the Gelb effect would still be a failure of color constancy. Misrepresentation occurs, according to the Bayesian model, when the illuminant falls outside of the parameter specified by the prior, or when the ‘cues’ used to recover the illuminant are not reliably correlated with the illuminant (i.e. ‘invalid’). In both conditions, the most likely representation given the stimulus is *not* the actual stimulus. Misrepresentation is normal, systematic and completely predictable.

Successful representation is distinguished from unsuccessful representation not in terms of the processes of the visual system, but rather in the information the visual system uses to form its representation of the stimulus.\textsuperscript{36} In both cases, the visual system produces a ‘best guess’ given the information available. If that information *in fact* does not correlate with the property represented, misrepresentation occurs. If it *in fact* does correlate with the property represented, successful representation occurs. Thus, we can say that a state of the visual system S misrepresents a property P if and only if the information which makes S the most likely representation for that visual system of P does *not* *in fact* correlate with P. And a state of the visual system S represents a property P if and only if the information which makes S the most likely representation of P for that
visual system does in fact correlate with P. Note that I am not here defining ‘representation’, but rather distinguishing between successful, veridical representation and unsuccessful, ersatz representation.

The resultant theory of misrepresentation is attractive in a number of ways. First, it has all the power of representationalism to explain the failures of our visual system without the implicitly normative claims of ‘normal’, ‘optimal’, or ‘paribus’ that infect other representational approaches. Second, it is broadly externalist: veridical and ersatz representation are distinguished by the validity of information used by the visual system. That information, however, is outside the skull and unavailable to reflective consciousness. When confined to a priori reflection, we are unaware of whether our visual system represents veridically or not in any given case.

It may be objected that I have not said where the properties a given object is misrepresented as having come from. Suppose I see a black tinged with green when placed next to a bright red in full sun. The resultant apparent ‘greenish-black’ is a misrepresentation, easily explained by assuming that the visual system incorporates the color signal of the red patch into its representation of the illuminant, which when discounted from the color signal of the black patch yielding a slightly greenish representation of black. But where, one objects, does the greenness come from? This is an odd question, despite the frequency with which it is repeated in Philosophy. Consider the Müller-Lyer illusion. The equivalent question would be: ‘where does the longness of the apparently longer line come from?’ Or suppose that I hallucinate a pink elephant. Where does the elephantness come from? These questions are nonsensical. Berkeley is right: if this objection holds for color, it holds for shape and size. And contrapositively: if
it fails to hold for shape and size, it fails to hold for color. Finally, it should be noted that this is not a challenge to my particular brand of representationalism, but representationalist theories of perception as a whole.

Second, one may object here that the move to banish normative constraints from the analysis of perception would leave the priors unconstrained. How could one determine if the information is correlated in fact with the property represented if one does not previously know the correctness conditions for the representation of that property? An answer appears simple: the visual system we share has been selected for in an environment driven by a pervasive illuminant constraint: sunlight. Asking ‘why are the priors used in the Bayesian model set to the spectral composition of sunlight dawn-dusk shade-full sun?’ is akin to asking ‘why are the beaks of the Darwinian finches set to the food sources available?’ Having a visual system capable of producing stable representations of objects in precisely that variety of illuminants found in its world would mean great benefit for that organism.

But that is not the end of the story. It is my hypothesis that the priors can be adjusted or manipulated over the course of a given individual’s life by experience, conscious interference or simple development. Again, there is obvious benefit to the organism: if a particular individual lives above the Artic circle, its visual system may have to adjust to a different class of illuminants from individuals of the same species that live on the Tropic of Cancer. The priors that operate on the spectral reflectances may be phylogenetically fixed, or they may be adapted via experience—in both cases, I take it that this question of innateness versus learning is a question for Psychology, not Philosophy. And, technically, it is tangential to the core of my discussion here. I believe
it is enough to say that the initial set of priors is likely phylogenetically fixed by natural selection. They are modified through continual growth, reflection and experience, resulting in variation in the small scale while explaining massive agreement in the large scale. The strength of the theory lies in its breadth, not in its origin story. The theory simultaneously explains why we agree on color appearance most of the time, suffer from the same massive constancy failures, yet vary when it comes to any given particular color appearance. That is a remarkable achievement indeed.

4. Conclusion

Color constancy fails. But it does so in a normal, systematic, predictable manner. The categorical view lacks a coherent explanation of this systematic failure. According to Thompson’s theory, agency is an integral part of the mechanism of color constancy and color appearance. Agents are, by definition, unpredictable. Therefore, we cannot successfully predict how a given stimulus will appear to an individual in a given scenario. Thus, there is no apparent ability to predict the systematic failure of color constancy.

The alternative view given in the literature, which I call ‘simple representationalism’, suffers from two well-known difficulties. First, it fails to fully explain how visual representations really are open to influence by consciousness; and second, it fails to provide a complete account of misrepresentation, relying merely on non-optimal conditions, ceteris paribus conditions or abnormal causal histories.

There are no such problems for Bayesian approach. By building on the representationalist tradition of discounting the illuminant, it is able to explain precisely when, and in what quality, misrepresentation will occur. And by including the Bayesian
priors, it allows for precision in specifying the role of expectation and personal histories in perception, thus accounting for inter- and intra-personal variation.

**Works Cited**


Notes

1 Thanks to the anonymous reviewers of Philosophical Psychology for their helpful comments.

2 This definition follows that used by Kraft, Maloney, & Brainard: “the ability of the visual system to process the retinal image and produce a perceptual representation of surfaces that is stable against variation in the illumination” (2002, p. 247).

3 Maloney & Schirillo define constancy success: “The human visual system assigns colors to surfaces. If the assigned color of a surface patch is (approximately) determined by the spectral properties of the patch itself, then the visual system is (approximately) color constant” (2002, p. 135). While this definition is attractive, it appears to be committed to something like the primary quality view of color. For the purposes of this paper, it is useful to have a notion of successful color constancy that does not require such a commitment

4 Kraft & Brainard state, for the purposes of their experiments with achromatic constancy: “good constancy is indicated when the shift in achromatic chromaticity has approximately the same direction and magnitude as the shift in illuminant chromaticity” (1999, p. 309).

5 “We having, by use, been accustomed to perceive … …what alterations are made in the reflections of light by the difference of sensible figures of bodies; the judgment presently, by habitual custom, alters the appearance into their causes. So that which is truly a variety of shadow… the perception of a uniform color, when the idea one receives from thence is only a plane variously colored, as is evident in the painting” (Locke, 1975, Bk I, Ch IX, §8)

6 It is also worth noting that Samuel Johnson made the implications of this thesis one of the grounds for criticism in his famous correspondence: “…Nor can I see how it follows that there is no external absolute height, bigness, or distance of things, because they appear greater or less to us according as we are nearer or remote from them…” (1998, p. 169).


Again, this line of argumentation is not new: Locke’s distinction between the primary / secondary qualities made in the Essays (Book 2, Ch. VIII, §9-10) is merely a distinction. His argument for the thesis that colors are secondary qualities appears in two versions in the Essays, an argument from porphyry, which is seen as
red and white in the light, but colorless in the dark (Book 2, Ch. VIII, §19) and an argument from the different appearance of blood seen under a microscope (Book 2, Ch XXIII, §11). Hardin, in his book Color for Philosophers, pointed to the fact that colors contrast with one another spatially, while primary qualities do not appear to do so. Johnston has famously asserted that colors are ‘fully revealed’ in perceptual experience, while primary qualities are not. (1996). And Thompson, Palacios & Verela have argued that colors have a unitary / binary structure that no other primary quality could possibly have (1992).

While I do not wish to put words into their mouths, the view I have in mind here is articulated by Wright in his 2003 and Wright and Johnson in their 2006.

Of course, the possibility of an unconscious was denied in Berkeley’s time.

Judd (1960, p.258) claims that as Helmholtz believed these ‘inferences’ to be automatic and unconscious, the distinction between principle and modern representational views is merely an equivocation on ‘judgment’. I am not so sure. Helmholtz explicitly maintains that these inferences are ‘learned’. However, if one was committed to the position, one could argue that the difference is merely the distinction between innate and learned concepts I dismiss as tangential in the next section.

See Katz, 1935, p. 14-15 for his discussion of this change in theoretical viewpoint. In recent years, Katz’s term ‘articulation’ has come to denote whatever attributes of the visual field enable color constancy’s success. Candidates that have been suggested include the numerosity of objects (Koffka, 1935, Linnell & Forster, 2002)¹¹, variations of colors in the visual field (Mausfeld & Andres, 2002)¹¹, spatial location (Schirillo & Shevell, 2002)¹¹, the three-dimensional structure of the space (Katz 1935, Bloj & Hurlbert, 2002)¹¹ and finally what Kraft, Maloney & Brainard call ‘valid’ cues (Kraft et al., 2002).

Witness this comment in his classic Katz, 1950 Gestalt Psychology:

All visual precepts are influenced by knowledge that comes with experience. But experience by no means plays the major role in forming objects into separate entities. Objects constitute themselves for other, more deep-seated reasons, and it is these which account for our ability to have experience with objects in the first place...

According to Gestalt Psychology, as opposed to the older theory, the tendency to form objects would become operative in the consciousness of a child from the start, even without experience of any sort... ...we often comprehend objects as units before we have any way of knowing what they are like (p 22).

Galen Strawson argues in his 1989 that linguistic agreement is necessary to determine if two individuals are seeing ‘the same’ color or not. Non-linguistic creatures, then, cannot be said to be seeing ‘the same’ color as we, or even they, see. If we extrapolate from interpersonal similarity judgments to intra-personal similarity judgments, it would follow that non-linguistic creatures cannot be said to be seeing ‘the same’ color on an object as they saw in different lighting conditions.

The reader will no doubt notice the similarity between this argument and the private language argument. I do not wish to run afoul of a body of literature on that argument, but this formulation appears amenable to empirical test. As I mentioned in the last section, experimental psychology has shown repeatedly — since the 1930s — that non-linguistic creatures as simple as goldfish are color constant. If the private language argument extrapolated above entails that they are not color constant, there is something wrong with the either the private language argument or its extrapolation.

Here is the relevant passage: “Useful vision depends on perceptual representations that make explicit the properties of the environment (i.e. the distal stimulus) rather than properties of the retinal image (i.e. the proximal stimulus). In the case of color, the distal stimulus may be taken as the spectral reflectance of the objects...” (Kraft, Maloney, & Brainard, 2002, p. 247).

The term ‘memory color’ has a different meaning today. That meaning, incidentally, was coined by David Katz in his criticism of Hering’s use of the phenomenon to explain color constancy. See Hering, 1964, p.7-8. Katz’s experiments are initially reported in Katz, 1911; described in Koffka, 1935 (p. 241) and Katz, 1935, (p. 162).

I am not suggesting here that Noë holds that constancy is determined by judgment. He clearly distinguishes his sensory-motor approach to perceptual content from those who find all perception to be conceptual in Ch. 6 of his 2004. But so did Helmholtz. Helmholtz initially proposed conceptual judgment as part of the mechanism of constancy, but later, in the face of empirical evidence changed the notion to ‘unconscious inference’. Noë denies judgment’s role in color constancy, but builds his theory on
'sensorimotor understanding', which is gained 'unconsciously in childhood'. (e.g. 2004 p. 183) The critiques leveled against Helmholtz are still applicable.

In section 6.2 of his 2004 *Action in Perception*, Noë tells us he finds the argument from infants and animals unconvincing because “there is a way of bringing concepts to bear in thought and reasoning other than their deployment in deliberative judgment” (p. 186). I agree. Indeed, it is why I list the distinction between ‘innate’ and ‘learned’ concepts as tangential in section 1.3. However, Noë is committed to the idea that concepts, analyzed as sensorimotor understanding, are deployed in every act of constant color perception; and hence, must be deployed in constant color perception of animals and infants. If Noë is willing to allow for possession of such concepts by infants and animals, as Thompson does, the evidence presented in this section is irrelevant to his view. If those concepts are gained unconsciously in childhood, we would expect non-constant representation before that process completes, and his view would be subjected to empirical critique on the basis of lack of evidence.

For Noë, the issue becomes articulating the mechanism of the acquisition of sensory-motor understanding, Given these further wrinkles in the dialectic, which I take to be empirical in nature and therefore tangential to this discussion, I will not address Noë’s position directly in section 2.3, but rather only the position that assumes a view of concept possession that precludes infants and animals.

Arend & Reeves’ experiment has been subject to a great deal of criticism in the literature, see, Brainard & Freeman, 1997, Kraft et al., 2002 and Delahunt 2001.

Many commentators have questioned the legitimacy of the task description, but Kuriki and Unikawa (1996) and Bauml (1996) both confirm that these tasks can be understood by naïve observers; although Bauml does explicate the tasks slightly differently, by claiming that the paper-match task involved a relational judgment while the hue-match task did not.

I am indebted to one of the anonymous reviewers of *Philosophical Psychology* for this point.

Katz continues, “On the other hand, I should not dare to assert *a priori* that the child sees in exactly the same way as an adult sees. In the pre-linguistic period this *could not be true*, because during this period *colours are not apprehended as belonging to specific categories*, and the significance of *categorical apprehension for colour-constancy has already been emphasized*. A further assumption might be that the child has from the beginning the same kind of object-experience as the adult. This is improbable, at least during the period before the child has learned to manipulate objects with his hands.” (Katz, 257, italics in the original)

It’s worth noting that even though I am appropriating Dannemiller’s work to provide a possible argument for the categorical view, Dannemiller himself hold a computational view of color constancy which holds, like those in the tradition of Maloney & Wandell, that task of the visual system is to recover the surface spectral reflectance (Dannemiller, 1989a).

Suttle, Banks & Graf report that infants red-green opponent-processing channel is functioning at 2-3 months, but the blue-yellow one only appears at four months (2002), while Adams & Courage report that prior to 10-11 weeks, infants fail in both Long- and Middle-wavelength discrimination tasks (2002).

This result was originally mentioned by Katz in his 1935, but has recently been collaborated with respect to color constancy tests by de Fez, Capilla, Luque, Perez-Carpinell, & del Pozo, 2001.

Buchsbaum’s 1980 paper, which continued the tradition of estimating the illuminant from the average reflectance of the scene was, obviously, prior to Maloney and Wandell’s 1986. For whatever reason, Buchsbaum’s work did not spark the interest that Maloney & Wandell’s did. Since 1986, multiple models have been proposed, new techniques have been investigated, and there are now 4-5 major papers on color constancy published annually. This just wasn’t the case between 1980 and 1986.

This criterion for adequacy of color constancy models was recently proposed by Alan Gilchrist (Cataliotti & Gilchrist, 1995; Gilchrist, 2003; Gilchrist et al., 1999). It is not surprising, to me at least, that he is one of the few experimental psychologists who have a fine appreciation for the work of the Gestalts.

Models of this sort, sometimes called ‘equivalent illuminant’ models have been proposed by Katz, 1911 and 1935, Beck, 1959, Maloney & Yang, 2003; Rutherford & Brainard, 2002, among others.

The terms used here are Brainard’s, but there is only a slight difference between this formula and the ‘Brunswik ratio’, proposed by Brunswik in 1928. The Brunswik ratio used the reflectances themselves, rather than the color signals under standard illumination. The change is necessary to allow for metamer matching. The Thouless ratio (Thouless, 1931) used the logarithms of the reflectance. For a discussion, see Beck, 1972 Ch. 4.
There appears to be a bit of confusion in the informal, i.e. non-technical, literature regarding the limits of color constancy. David Brainard and colleagues have shown that in ‘nearly natural viewing conditions’, color constancy is remarkably robust. However, Helson & Judd (1936) are often cited as having shown that color constancy can fail quite dramatically when the background conditions vary (see Maloney & Schirillo, 2002 for an example). This is a little odd, as Judd himself conjectured in 1940 that the human visual system is exactly color constant for illuminants within the range of natural sunlight. The confusion around these issues appears to be regarding the fact that Helson & Judd believed color constancy and color contrast to be phenomena produced by a single mechanism. Thus, they distinguish between two broad classes of errors in what they called ‘color conversion’: illumination-dependent errors and background-dependent errors. The former are what we now call ‘failures of color constancy’, while the latter are usually called ‘color contrast effects’. While there must be a connection between color contrast and color constancy, I do not wish to take any position on this debate. Suffice it to say that where, and in what way, color constancy fails is vitally important matter to both the categorical and representational explanation.

Recently, after seeing an exhibition of Rob & Nick Carter’s work ‘Painting with Light’ in London, Spehar & Clifford called for more investigation into the exact nature of failures of color constancy in the pages of Perception. At the very least, I think everyone working in this field would agree (2002).

In an unpublished work, Jonathan Cohen suggests that the phenomenon of color constancy has been widely misunderstood. Using the ‘equivalent illuminant’ model of David Brainard, he argues that asking a subject to match a colored patch under variant lighting conditions is equivalent to the counterfactual: ‘Would these two patches match if (counterfactually) they were presented under the same illumination conditions’. But the counterfactual approach is subject to Occam’s razor. In order to represent the light reflected by an object in a different illumination condition, the visual system would first have to calculate the spectral reflectance of the object.

Kraft et al. do state, “The equivalent surface would have appeared achromatic had it been placed at the test patch location with the projection colorimeter turned off” (Kraft et al., 2002, p. 10), but they do not claim the counterfactual view. According to Brainard, the equivalent surface is the spectral reflectance that is represented by the visual system. If Cohen’s view is correct, the visual system would have to do a further calculation: the equivalent spectral reflectance multiplied by some standard illumination. Why the extra step? In order to make the counterfactual view work, the visual system must complete the phenomenal representation, and then go further. The further step is unnecessary.

Moreover, these experiments only seek to establish that the visual system is, indeed, discounting the illuminant to recover the spectral reflectance. It makes no commitment to how of representation — specifically the mechanism by which it completes that task. That is the interesting part of color constancy research today, and his model of that mechanism is what sets his research apart from all of the other work that has been completed in the representationalist tradition.

According to Brainard, the equivalent illuminant model treats “the visual system’s estimate of the illuminant as a parameter, so that we test the idea that the visual system estimates an illuminant without committing ourselves to how this estimate is obtained” (1997, p. 2107, italics mine). Bühler, for example, appears to have theorized that the illuminant is recovered from the reflectance of dust particles floating in the air (Katz, 1935, p. 268-9). This view was discarded long ago, but it is compatible with the equivalent illuminant model.

The Gestalt’s theories, specifically Koffka’s theory of ‘foreign fields of illumination’, are one kind of classical model. I will not discuss these in depth in this paper, as they are not considered viable in contemporary research.

Those versed in the history of Gestalt psychology may recognize the resemblance of this view with Koffka’s ‘law of prägnanz.’ Of course, this view is significantly more precise in how it calculates what is considered the ‘simplest’ explanation of the given stimuli.

Unfortunately, the paper is still under review, and I cannot, at this time, quote from or cite it. It is available, however, from David Brainard’s web site. There is little doubt in my mind that by the time this paper is reviewed, that paper will be published, and I can change this sentence to give some more specifics.

In 1939, Duncker reported that qualitative matches of grey leaf-shaped test patches under a reddish light were closer (for the naïve participants in his experiment) to green than donkey-shaped grey test patches in the same experimental conditions. The shifts in color appearance were minor and sophisticated participants—a painter, psychologist and the experimenter—did not demonstrate the memory color influence. This lead Duncker to conclude that memory color may influence hue matching, and needs to be
studied more closely, but probably is not the operative mechanism of color constancy. Following Duncker, Bruner, Postman and Rodriguez showed that desaturated photographs of tomatoes and lemons were judged more saturated under conditions of color contrast than ovals or squares (1951). Ultimately Bruner and Postman went on to their famous anomalous playing cards demonstration, which demonstrates— that is, if one sees the anomalous red spade as black—a failure of memory color, not a failure of color constancy.

36 Let me say outright that I do not have a worked-out theory of information behind this claim. Intuitively, information carried by light is those properties of the light that causally co-vary with the properties of the object the information is about. If this notion of information threatens to push my view closer to that of the representationalists I have criticized, so be it. My dispute with them is not over the core of the representationalist doctrine, but rather on their inability or unwillingness to explain normal variation and constancy failure.