Colour Constancy as Counterfactual

Jonathan Cohen

There is nothing in this world constant, but inconstancy.
— Jonathan Swift, "A Critical Essay upon the Faculties of the Mind".

Abstract

In this paper I argue that two standard characterizations of colour constancy are inadequate to the phenomenon. This inadequacy matters, since, I contend, philosophical appeals to colour constancy as a way of motivating illumination-independent conceptions of colour turn crucially on the shortcomings of these characterizations. After critically reviewing the standard characterizations, I provide a novel *counterfactualist* understanding of colour constancy, argue that it avoids difficulties of its traditional rivals, and defend it from objections. Finally, I show why, on this improved understanding, colour constancy does not have the philosophical consequences that have been claimed for it in the literature.

In recent years, vision scientists and philosophers of perception have devoted considerable attention to questions about colour constancy. Among the most important issues surrounding colour constancy are those about how we should understand the phenomenon and those about what the phenomenon shows about the nature of colour. Many writers explicitly link these two sets of questions: they have enlisted certain kinds of answers to the former set as a way of arguing for certain kinds of answers to the latter set. In particular, many have appealed to a certain understanding of the empirical phenomenon as evidence for the view that colours are illumination-independent: viz., many have taken constancy as support for the widely-held position that colours are subject-independent, physical properties of objects.[1] All the vocal proponents of this view (e.g., [Hilbert, 1987], [Byrne and Hilbert, 1997a], and [Tye, 2000]) have taken colour constancy as a central pillar of support, and even those who reject the view typically allow that colour constancy is something they need to discount or play down in order to meet their opposition (e.g., [Hardin, 1988], 46ff; [Hurvich, 1981], 199; [Jameson and Hurvich, 1989]).

I believe that the phenomenon of colour constancy has been misunderstood, and that correcting this misunderstanding will have the effect of undercutting the apparent support the phenomenon provides to accounts of colour as

---

[1]Among such views (which have sometimes gone by the names 'physicalist', 'primary quality', or 'objectivist' theories of colour), the most currently influential position holds that (surface) colours are identical to (classes of) surface spectral reflectance distributions.

illumination-independent properties.[2] Aside from its independent interest, the exercise of correcting these faults will also provide a methodological moral about the way in which arguments over colour ontology should proceed: it will remind us to remain on guard against characterizing phenomena in ways that tacitly beg disputed metaphysical questions.

In §§1–2 I'll examine critically some attempts to characterize the phenomenon of colour constancy. Next, in §3, I'll consider the ontological lessons some have attempted to draw from the phenomenon. Finally, in §§4–5, I'll offer my own characterization of colour constancy, argue that it does not suffer from the problems that plague other characterizations, defend it from objections, and show why, on this improved understanding, colour constancy does not have the philosophical consequences that have been claimed for it in the literature.

1   Constancy as Invariance

Since colour constancy is easier to recognize than to define, I'll begin with the example depicted in figure 1 — a coffee cup on a table, partially in direct sunlight and partially in shadow.[3] Consider the region of the coffee cup (and the region of the table) in direct sunlight, and compare it against a similar-sized, adjacent region of the coffee cup (and a region of the table) that are in shadow. If you are like most normally sighted subjects, you will find that these two regions are, in some sense to be explained, alike in apparent colour.[4] On the other hand, again

---

[2]I suspect that one source of misunderstanding has been the tendency by philosophers writing about colour and colour constancy to focus on a limited portion of the relevant empirical literatures. This is true not only of the psychophysical literature on colour constancy to which I'll advert below, but also of computational accounts of colour constancy. For example, the discussions of computational theories in [Hilbert, 1987] and [Byrne and Hilbert, 2003] emphasize only theories that model the recovery of (illumination-independent) spectral reflectance distributions from the energy falling on the retina (e.g., [Maloney, 1986], [Wandell, 1989]). But this overlooks other computational models of colour constancy that attempt to determine whether two regions are relevantly alike (though different in illumination) without extracting an illumination-independent representation of the regions (e.g., [Craven and Foster, 1992], [Zaidi, 2001]).

[3] The case I have in mind is one in which the subject is visually perceiving a coffee cup, not a photograph of the coffee cup (notwithstanding the arguments of [Walton, 1984], I take it that when a subject sees a photograph of x, she does not ordinarily see x as well [Cohen and Meskin, 2004]). Hence, I am using figure 1 to *depict* the stimulus – I am not using it *as* the stimulus. This will matter, since it is plausible that there is a variation in the illumination under which the coffee cup is perceived, whereas it is doubtful that there is a significant variation in the illumination under which figure 1 is perceived. For ease of expression, I'll be pretending that the reader can perceive the depicted coffee cup in what follows. (*Mutatis mutandis* for depicted objects in other photographs throughout.)

[4]In talking of apparent colours of objects, I mean the colours that objects are represented by subjects (/visual systems of subjects) as having. I see no reason to regard the recognition of apparent colours (as properties over and above the colours that objects have) as ontologically profligate. It is a banal fact about objects that they exemplify colours, sizes, shapes, smells, and many other properties. It is a banal fact about us that we represent objects as exemplifying colours, sizes, shapes, smells, and many other properties. But the properties we represent objects as having can't be identified with the properties objects have, since we sometimes misrepresent the properties of objects.

Figure 1: Partially sunlit coffee cup.

assuming you are like most normally sighted subjects, you will also find that the regions are easily, obviously, and quickly visually discriminable in apparent colour (in some sense also to be explained). These two reactions to a pair of visual stimuli (in this case, the pair consists of the two simultaneously presented regions of the coffee cup) are the hallmarks of colour constancy.

Now, the case just presented is an instance of *simultaneous* colour constancy in that the two patches under consideration are presented simultaneously. But we can also consider cases of *successive* colour constancy, in which a test and comparison patch are presented successively/non-simultaneously.[5] The photographs in figure 2 depict a set of common objects taken under different daylight illuminants.[6] If we perceive each photograph successively, we can ask about the apparent colour of some object (say, the banana) that is represented in each

---

[5]Despite the common nomenclature, we should be cautious in assimilating simultaneous and successive colour constancy, as the relationship between the two is not presently well understood. It is likely that both simultaneous and successive colour constancy are interaction effects of a number of different mechanisms (operating over different time scales), and it is not clear whether or to what extent the same mechanisms are at work in the two phenomena. On the other hand, it is worth noting that psychophysical experiments on simultaneous colour constancy in fact involve successive perception of simultaneously presented patches: in such experiments subjects are always instructed to fixate alternately on one and then the other of the two simultaneously presented patches at 1 second intervals in order to avoid becoming adapted to either.

It should also be pointed out that figures 1–2 are not presented here under controlled psychophysical circumstances, and should be taken as merely illustrative of the phenomena of simultaneous and successive colour constancy.

[6]I am grateful to David Hilbert for supplying and allowing me to use these photographs.

of them. Although the effect is less obvious here (perhaps partly because comparison between non-simultaneous presentations introduces confounding considerations about perceptual memory), in cases of successive colour constancy we find the same pair of perceptual reactions that occur in cases of simultaneous colour constancy. On the one hand, normally sighted subjects find that the two (successively presented) regions of interest are, in some sense to be explained, alike in apparent colour. And on the other hand, normally sighted subjects find that the two (successively presented) regions of interest are, in some sense to be explained, easily, obviously, and quickly visually discriminable in apparent colour.

Many writers who have considered the phenomenon of colour constancy have emphasized two facts: first, that subjects classify the two patches under comparison as (in some sense to be explained) alike in apparent colour — this is the first reaction mentioned above; and second, that the two regions differ in the illumination falling on them. And this has led to a more or less standard understanding of colour constancy as a kind of invariance. In particular, on this view (henceforth, invariantism), colour constancy is an invariance of apparent colour across changes in illumination. Invariantism has become the *de facto* standard understanding of colour constancy in both philosophical and scientific work on colour; for example, versions of this characterization can be found in many recent textbooks and anthologies on colour and vision (often in glossary entries for 'colour constancy') including ([Byrne and Hilbert, 1997b], 445), ([Zaidi, 1999], 339), ([Goldstein, 1999], 567), ([Stoerig, 1998], 141), and ([Brainard et al., 2003], 308–309). While these formulations differ slightly in their details,[7] they share the core idea that colour constancy should be regarded as an invariance in our perceptual reaction to members of a pair of stimuli despite differences in the illumination under which each member of the pair is viewed.

Moreover, as I've said, the invariantist understanding of colour constancy has played an important role in arguing for views according to which colours are illumination-independent properties of surfaces. To a first approximation, the argument is that, if apparent colour is indeed invariant across differences in illumination, then the theory that best explains that fact is one that takes colours to be illumination-independent properties.[8] (I'll return to consider this

---

[7]In particular, they disagree about whether what is invariant across changes in illumination in cases of colour constancy is "perceived color" (Byrne and Hilbert) "the color percepts assigned to individual objects" (Zaidi), "the perception of an object's hue" (Goldstein), and so on; needless to say, a more thorough search of the literature would turn up yet other formulations. For expository convenience I'll be concentrating on the idea of invariance of apparent colour across changes in illumination. As far as I can see, nothing of significance hangs on this particular choice: the points I'll be making can be extended easily to the other formulations of invariantism.

[8] That the phenomenon plays this role in motivating illumination-independent accounts of colour also explains why it is in the interest of the invariantist to describe the phenomenon as an invariance of *apparent colour* rather than an invariance of *colour*. Namely, since the assumption that the two regions of interest in figure 1 share a colour is tantamount to the (disputed) conclusion that colours are illumination-independent, building that conclusion into the description of the phenomenon would deprive the latter of any capacity to provide indepen-

Figure 2: Common objects under different daylight illuminants.

argument in greater detail in §3.)

Since the invariantist conception of colour constancy plays such an central role in motivating theories about colour ontology, it is worth asking how well that conception is supported by the facts.

I want to argue that invariantism is threatened by something already noted in our initial comments about figures 1–2: while there is some sense in which a subject's reactions reveal the apparent colours of the regions to be alike, there is also some other sense in which the same subject's reactions reveal the apparent colours of the regions to be unalike. This fact is, I want to suggest, a serious problem for invariantist accounts of colour constancy. For, if the foregoing is correct, then it is at best seriously misleading to say, with the invariantist, that colour constancy involves an invariance of apparent colour across variations in illumination. This claim respects only one of the two reactions normally sighted subjects have to these cases. While it respects one of their reactions — their reaction that the two members of the pair of stimuli are alike in apparent colour, it ignores their reaction to the effect that the members of the pair *differ* in apparent colour — that there is *not* an invariance of apparent colour across changes in the illumination.[9] For the sake of convenience, call the first reaction 'the invariance reaction', and the second reaction 'the variance reaction'. The problem for the invariantist about colour constancy is that her characterization of the phenomenon is consistent with the invariance reaction, but not the variance reaction.

In fact, the problem I'm pressing can be expressed more precisely by turning from the evidence of qualitative discriminations that I've been relying on so far to the sharper, quantitative measurements used in contemporary psychophysical investigations of colour constancy. The main quantitative measure by which contemporary psychophysicists assess colour constancy, known as asymmetric colour matching ([Wyszecki and Stiles, 1982], 281–293), involves asking subjects to change the chromaticity (or lightness, in lightness constancy experiments) of a test patch under one illuminant until it perceptually matches a standard patch under a different illuminant. Since the subject eventually arrives at a perceptual match, investigators regard the chromaticity (/lightness) difference between the test and the standard patches as a quantitative measure of the perceptual effect

---

dent support for the conclusion. In contrast, if the invariantist characterizes colour constancy as an invariance in apparent colour, we can take her as arguing for a theory according to which the two regions share a colour on the *prima facie* rational, non-question-begging, ground that that theory sustains naïve appearances.

[9] I am not the first to notice this point, or to conclude from it that invariantist characterizations of colour constancy are inadequate; see, for example, [Craven and Foster, 1992], 1360; [Noë, 2006], §3 and [Noë, 2004], chapter 4; [Thompson, 2006], 79ff. (Noë, in particular, responds to the difficulty along something like the lines of the proposal I'll make in §4 by taking the visual system to be responsive to counterfactual properties of objects in the visual scene. But my view differs from Noë's in two significant respects: (i) Noë construes the antecedents and consequents of the relevant counterfactuals as being much more tied up with action than I do; and (ii), for him this sort of counterfactual dependence is part of a much broader conception of perception as tied to sensorimotor counterfactuals, whereas I hold that the counterfactual dependence involved in colour perception is a relatively isolated case.)

of the illumination difference.[10]

How can asymmetric colour matching help us to assess invariantism about colour constancy? According to invariantism, colour constancy is a kind of invariance across illumination differences; hence, on this view, colour constancy will be exhibited to the extent that the effects of illumination differences (as measured by chromaticity/lightness differences in such experiments) are small. So what do the results of asymmetric colour matching experiments show about colour constancy?

Unsurprisingly, asymmetric colour matching brings out the very same points considered above, and does so more precisely than the qualitative measure discussed. Roughly, it seems that (most) subjects can respond in two different modes — one mode that assimilates surface pairs like that in figures 1–2, and one that distinguishes such pairs. Moreover, it seems that (most) subjects can be made to switch between these two modes of response as a result of experimental instructions: instructions to "adjust the test patch to match its hue and saturation to those of the standard patch" lead subjects to distinguish members of such pairs, while instructions to "adjust the test patch to look as if it were 'cut from the same piece of paper' as the standard, i.e., to match its surface color" ([Arend and Reeves, 1986], 1744) lead subjects to assimilate members of such pairs.[11] (I'll follow [Bäuml, 1999] in talking about the first sort of matches — those that distinguish the two surfaces — as *appearance matches*, and the second sort of matches — those that assimilate the two surfaces — as *surface matches*. Obviously, no substantive conclusions should be inferred from the choice of these labels.)

---

[10]Two methodological remarks are in order.

First, until recently, most psychophysicists using asymmetric colour matching to study colour constancy have in fact used colour contrast as a proxy for changes in illumination. That is, the typical experimental display consists of a test and a standard patch set against circular surrounds of differing chromaticity, but with the whole scene under uniform illumination (indeed, usually surface illumination is negligible, as such displays are produced on luminous computer monitors). Subjects are then asked to change the chromaticity of the test patch against its surround until it perceptually matches the standard patch against its surround. On its face, this task seems to use chromaticity changes to measure the effect of a difference in surround, rather than the effect of a difference in illumination. Fortunately, a number of investigators (led by David Brainard and his colleagues), have begun to measure the effects of illumination differences more directly (see, for example, [Brainard et al., 2003]). In correspondence, Brainard has indicated to me that the patterns of results that I'll be presenting below are sustained in the more direct experiments he employs. As such, I'll be ignoring this complication in what follows.

Second, for obvious reasons, the techniques under discussion here are not applicable to the study of colour constancy in non-human and infant human subjects (see §4.2), so investigators have had to rely on other methods in these contexts.

[11]These results have been confirmed by a number of studies. See [Evans, 1948], 163–164 and [Beck, 1972], 66–67 for an overview of some of the earlier work; more recent findings to the same effect (typically restricted to simultaneous colour constancy) are reported in [Arend and Reeves, 1986], [Blackwell and Buchsbaum, 1988], [Valberg and Lange-Malecki, 1990], [Arend et al., 1991], [Troost and deWeert, 1991], [Cornelissen and Brenner, 1995], and [Bäuml, 1999]. While there has been far less systematic investigation of this effect with respect to cases of successive colour constancy, the bimodal pattern of results depending on instructions has been found for successive colour constancy as well ([Delahunt, 2001], 114–117; [Delahunt and Brainard, 2004], 71–74).

So, once again, the problem is that, while the invariantist characterization of colour constancy is consistent with one reaction that subjects have to canonical cases (as measured by qualitative and quantitative methods), it is inconsistent with another, equally good reaction that subjects have to the very same cases (as measured by qualitative and quantitative methods). Invariantism predicts the invariance/surface match reaction, but not the variance/appearance match reaction. What should the invariantist say about this situation?

1.1 Invariantist Responses

First, the invariantist might attempt to discount the problematic variance reactions by insisting that, when subjects have both variance and invariance reactions to a pair of surfaces, the case is not an instance of colour constancy. But this alternative strategy seems inadvisable. After all, insofar as the pair of conflicting responses seems to be a hallmark of paradigm cases of colour constancy, the strategy in question amounts to emptying the phenomenon of instances merely to save a favored theory. Presumably this is unacceptable.[12]

A more promising invariantist strategy would be to find an alternative characterization of the troublesome variance reaction as a discrimination along some dimension other than that of apparent colour. If the invariantist can do this, there would be no need to take the variance reaction as contravening the core invariantist idea that colour constancy involves an invariance in perceptual reaction despite differences in illumination. In particular, an obvious alternative characterization would be that the visual system is here discriminating a difference in the illumination falling on the surface regions (a difference that must exist, given the setup of the case) rather than a difference in apparent colour.[13]

Now, in order for this suggestion to constitute a genuine alternative to the claim that the variance reaction reveals a difference in apparent colour, the invariantist must be thinking of illumination as something that is not merely a dimension of apparent colour.[14] But there is reason for doubting that illumina-

---

[12]Similarly, an interlocutor might object that visual discriminations are unreliable in cases where there is a difference in the illumination falling on the two samples; therefore, she might say, the cases I've appealed to fall outside the realm within which visual discriminability is an adequate test of difference in apparent colour, and so should be put aside. But once again, this line of response seems unwise in the present setting; for if we are debarred from considering discriminations between regions that differ in illumination, then we can never decide whether regions are invariant in apparent colour across variations in illumination. But variations in illumination are precisely the variations that we must consider in assessing whether or not there is colour constancy, as that phenomenon is understood by an invariantist; therefore, the proposed strategy of setting aside subjects' visual discriminations in such cases would mean that these cases — paradigmatic cases, and therefore precisely the cases that an adequate understanding of colour constancy ought to capture — could never be regarded as instances of colour constancy (on an invariantist account).

[13]I'm putting aside obvious dimensions of visual discriminability such as size, form, or texture as possible explanations; for, even if there are salient differences along these dimensions in figures 1–2 in particular, there are other paradigm cases of colour constancy (which we can produce in the psychophysics laboratory as needed) in which these possible confounding sources of discriminability are eliminated.

[14]Besides constraining the interpretation of the invariantist proposal on the table, this con-

tion, understood in this way, can provide an adequate account of the variance reaction.

To see why, consider, once again, the appearance matches described above. When subjects make appearance matches in canonical cases of colour constancy (e.g., that depicted in figure 1) they make the regions cease to be discriminable (along whatever dimension they were previously discriminable) by adjusting the hue and saturation of one of them. Now, it is a standard assumption in visual psychophysics that the hue and saturation of a patch are dimensions of its apparent colour; if so, then adjusting the hue and saturation of the test patch just is adjusting the patch's apparent colour. Therefore, whatever the difference was in virtue of which the patches were initially visually discriminable, that difference can be offset by a difference in apparent colour. And this, in turn, might lead us to suspect that the difference revealed in the variance reaction *is* a difference in apparent colour, contrary to the invariantist strategy now under consideration.[15]

An invariantist might respond that, even if the discriminable difference can be offset by a difference in apparent colour, we need not conclude that it *is* a difference in apparent colour. Instead, she will explain the facts in the following way. She will say that, prior to the subject's manipulation, the regions initially differ in illumination, but that they share a common apparent colour. When the subject modifies the hue and saturation of one region, thereby changing its apparent colour, the invariantist will claim that this manipulation also has the effect of changing the illumination of that region until, at the end of the modification, the two regions are alike in illumination, but different in apparent colour. The variance reaction will cease at this point, she will say, because that reaction is driven by a difference in illumination (and not apparent colour), and the latter has been obliterated by the modification.

But the invariantist's redescription of the facts is implausible. For one thing, if, as the invariantist must insist (in order to avoid the conclusion that the

---

sideration also gives us reason for rejecting a hybrid view according to which what is discriminated in canonical cases is some combination of a difference in apparent colour and a difference in illumination. Such a hybrid view allows that the regions of interest (e.g., in figure 1) differ in apparent colour (*inter alia*), and so cannot sustain the invariantist commitment that the regions are alike in apparent colour.

[15]Similar considerations tell against a different proposal for defending invariantism (suggested by an anonymous referee) building from the observation that we might take talk of 'apparent colour' ('perceived colour', etc.) not as meaning 'colour the object is represented as having', but as something more like 'way the colour appears'. One might hope to save the invariantist by taking her core commitment as insisting that the regions are alike in the colour they are represented as having, even while allowing, on the basis of their visual discriminability, that the two differ in the way that that common colour appears in them. This proposal is like that considered in the main text in that it proposes to account for the variance reaction as a discrimination of something other than the colour the regions are represented as having — in this case, a discrimination in the way a (shared) colour appears. As with the proposal considered in the main text, this view runs into the difficulty that the difference between the regions is, as a matter of empirical fact, not independent of differences in hue and saturation; and if the latter are differences in the colour the individuals are represented as having, it follows that the discriminated difference between the regions is not independent of differences in the colour the regions are represented as having.

regions at issue differ in apparent colour), illumination is not a dimension of apparent colour, it is hard to see why a modification in the apparent colour of a region should also change its illumination. Moreover, it is simply untrue that two regions will cease to be discriminable when they are different in their apparent colour (hue, saturation, and lightness) and alike in the illumination falling on them; on the contrary, subjects have no difficulty discriminating regions under those conditions.

Perhaps the invariantist will attempt to account for the post-manipulation discriminability of the regions by suggesting that the two are, at the end of the manipulation, discriminable along some dimension *other* than that of illumination (which, on the present story, accounts for the variance reaction). That is, she might suggest that the regions are discriminable along the dimension of illumination before the manipulation, but discriminable along some different dimension after the manipulation — indeed, the obvious suggestion is that the post-manipulation discrimination is a discrimination in apparent colour. But this, too, is hard to accept. For one thing, there is no phenomenological or psychophysical evidence for thinking that visual systems are responding to a different dimension of difference before and after the manipulation, as the invariantist suggests. Moreover, what the invariantist proposes about the post-manipulation situation is implausible on its face. The reason the manipulation comes to an end is that the regions are no longer discriminable *tout court*, not that the regions are no longer discriminable along the dimension of illumination albeit discriminable along some other dimension. If the invariantist offers a description of the situation that makes the regions come out post-manipulation as discriminable (as we are imagining), then we have reason to reject that description no matter how she proposes to analyze the discrimination.

Pending answers to these concerns, I take it that the invariantist response we have been considering is unsatisfactory.[16]

---

[16] A quite different objection to the line I'm pressing concerns the relationship between colour constancy and other forms of perceptual constancy. As it happens, colour constancy is just one among many perceptual constancy phenomena standardly explained in invariantist terms — e.g., size constancy is characterized as an invariance of perceived size across changes in the distance between the subject and the test object, shape constancy as an invariance of perceived shape across changes in the position from which the subject views the stationary test object, etc. If the considerations adduced so far against invariantism about colour constancy are persuasive, one might think, they should be equally telling against these other forms of perceptual constancy as well. But that sort of revision might seem too radical.

There are a number of things to say about this worry. First, before we generalize too quickly, it should be emphasized that the problem we're considering for invariantism about colour constancy arises from the conflict between the commitment of the view and the empirical facts about apparent colour. As such, the applicability of the present worry to size and other constancy phenomena is an entirely empirical matter: there is such an analogous problem for invariantism about (say) size constancy just in case, as a matter of empirical fact, invariantism about apparent size conflicts with the data about apparent size.

Assume, however, as seems not implausible, that the empirical facts about apparent size are relevantly analogous to those about apparent colour: suppose that we can get subjects to treat two disks of the same diameter, viewed at different depths, as unlike in apparent size. In this case, I think we would be justified in concluding that there's a good sense in which the two disks are unlike in apparent size, hence that the size-invariantist's claim that the two disks

1.2 Summing Up

Here's where we are. We've seen that invariantists claim that, in the cases under consideration, the regions of interest are treated by visual systems as being alike in apparent colour. On the other hand, we've seen that, while there is a sense in which visual systems treat the members of the pair as alike in apparent colour, there is another sense in which they treat the members of the pair as unlike in apparent colour. Or, in still other words, while there is a sense in which the pair of surfaces share a common apparent colour, there is another sense in which the pair of surfaces vary in apparent colour. So the situation cannot be quite as the invariantist describes it.

In the face of these data, an invariantist could reasonably cling to her characterization of the phenomenon if she could provide some reason for taking the half of the data that accords with invariantism (viz., invariance reaction/surface match data) seriously while ignoring the half of the data that does not (viz., variance reaction/appearance match data). But I don't see what that reason would be, and I certainly don't see that one has been given; on the contrary, for all that has been said, both sets of matches have an equal claim to reflect the visual system's representations about whether there is a single apparent colour or not.[17] Thus, pending some independent reason for siding with surface matches, and against appearance matches, the most we can say is that the existence of the invariance in terms of which invariantism characterizes colour constancy is supported by at most half of the data.

On the other hand, while the data don't appear to provide a direct confirmation of invariantism, it should be admitted that the results can no more obviously be claimed as a direct refutation of the view. After all, it seems significant that the data show two distinct patterns of response, rather than a continuous range. Moreover, it is suggestive that most subjects can be made to switch between the two patterns of response by modifying the experimental

---

are *alike* in apparent size (despite a difference in depth) is inadequate as a characterization of the phenomenon. (Likewise for other perceptual constancies.)

However, I want to suggest, this conclusion lacks the ontological significance for size that it has in the case of colour, and so is far less pressing. This is because we don't need the phenomenon of size constancy to serve as evidence for the conclusion that size is independent of viewing depth. In particular, we have agreed upon standards for size that are constituted independently of viewing depth (these standards are called 'rulers'), and that give us reasons for believing that size is independent of viewing depth. Thus, while rejecting invariantism about size constancy might require a more careful characterization of the phenomenon, it would not undercut our reasons for taking size to be independent of viewing depth. In contrast, invariantism about colour constancy is the main reason for thinking that colours are independent of illumination (and is presented as such by proponents of that view); as I have argued elsewhere [Cohen, 2004], there are no standards for the colour of x that are constituted independently of the illumination under which x is viewed. Thus, undercutting invariantism about colour constancy leaves the claim that colours are illumination-independent wholly unsupported.

[17]Of course, there is a difference in the instructions that elicit each set of matches — "adjust the test patch to match its hue and saturation to those of the standard patch" versus "adjust the test patch to look as if it were "cut from the same piece of paper' as the standard patch." But it is hard to find in this instructional difference (or, as noted, in the different labels chosen by the experimenters) any good reason for favoring the surface match data over the appearance match data.

instructions. This might be taken to suggest that we are looking here at two effects rather than one.

Given this situation, it seems to me that the most reasonable invariantist response would be to acknowledge that there are two different effects, and offer invariantism as a partial story about only one of them — the invariance/surface match reaction. But this can't be the end of the story. For one thing, a proponent of such a modified invariantist view owes us a more restricted reformulation of her invariantism, so that it is clear exactly what she claims remains constant across changes in illumination (since, as we have seen, it is misleading with respect to half the data to say that apparent colour remains constant). Moreover, it is fair to ask, of such an invariantist, what accounts for the variance/appearance match reactions, and how the two sets of reactions are related. Finally, we should ask how well a suitably restricted invariantism supports the ontological views about colour that it has been enlisted to defend.

In §4 I'll offer just such a revised view and defend answers to the questions just asked. (I take it to be a merely verbal matter whether a view resulting from such modifications should be counted a version of invariantism.) Before I come to this, however, I want to consider another alternative to invariantism (§2), and then reconsider the arguments for an illumination-independent conception of colour properties that turn on the phenomenon of colour constancy (§3).

2 Constancy as Reflectance Perception

I have argued that invariantism about colour constancy — the *de facto* standard description of the phenomenon — is unsatisfactory. There is, however, an alternative characterization of colour constancy that occasionally shows up in the literature (though much less frequently than invariantism). After motivating and explaining this alternative account, I'll argue that it fares no better than invariantism.

For reasons that will become apparent immediately, I'll refer to the alternative proposal in question as the *reflectance perception account* of colour constancy. The reflectance perception account can be motivated by the thought that, if invariantism fails because there is a good sense in which apparent colour is not (as a matter of empirical fact) invariant across changes in illumination, then perhaps we should characterize colour constancy in terms of something that we know, antecedently (as a matter of definition), to be invariant across changes in illumination.

As it happens, there are properties of surfaces that are (mostly) invariant across changes in illumination, and that many theorists have argued are identical to the colours: surface spectral reflectance distributions. Reflectance distributions are (certain exceptions aside) illumination-independent by definition.[18]

---

[18] Photographic (photosensitive) papers constitute one well-known class of exceptions to this generalization: this is because their dispositions to affect light are modified by exposure to light. The spectral reflectance distributions of photographic papers, then, are examples of the troublesome properties that some philosophers have called "finkish dispositions" —

Although a surface with a reflectance distribution R may reflect one way under illuminant I_1 and a different way under illuminant I_2 (and consequently may be thought to change its apparent colour under varying illumination), it will continue to have reflectance distribution R despite the change in illuminant. This result is unsurprising, insofar as reflectance distributions are dispositional properties — they are dispositions of surfaces to affect light of certain wavelengths in certain ways. Generally speaking, dispositions remain in place whether or not their activation conditions are satisfied: objects exemplify their dispositions before their activation conditions are satisfied, while their activation conditions are satisfied, and after their activation conditions are no longer satisfied. Table salt, for example, is soluble whether or not it is in fact immersed in water: it exemplifies the disposition to go into solution when immersed in water before it is ever immersed in water, at the moment that it is immersed in water, and after it is no longer immersed in water (say, by being precipitated out of water). Similarly, a surface that has the disposition to reflect 45% of the incident light of wavelength λ will (unless it is changed in some finkish way that would demand independent explanation, as in the case of photographic paper) exemplify that disposition before it is illuminated by light of wavelength λ, while it is illuminated by light of wavelength λ, and after it is no longer illuminated by light of wavelength λ. That is to say, the reflectance distribution of a surface is (exceptions aside) invariant across changes in the illumination under which the surface is presented.

Given that reflectance distributions are (exceptions aside) invariant across changes in illumination, and given that the trouble with invariantism is that, in one good sense, apparent colour is not invariant in this way, it is natural to turn to the reflectance account as a way of saving what is right about invariantism.

Formulations of colour constancy along these lines are advocated by a number of authors (although they are less prevalent than invariantist formulations):

> Color constancy is the ability to perceive the reflectance spectrum of surfaces despite changes in illumination and other viewing conditions" ([Palmer, 1999], 705).[19]

> The ability of a vision system to diminish, or in the ideal case, remove, the effect of the illumination, and therefore "see" the physical

---

dispositions that disappear when their manifestation conditions occur [Martin, 1994]. For obvious reasons, the reflectance perception account cannot be extended straightforwardly to these cases. I shall put aside this potential objection to the account, since I believe it faces more serious problems. (However, I'll return to the topic of finkish dispositions in defending my own account of colour constancy in §4.3.)

[19] Note that Palmer's mention of "other viewing conditions" suggests that he thinks variations other than those of illumination are relevant. Since this won't matter for our purposes, I'll put this matter aside. Also notice that Palmer's formulation describes constancy as involving our abilities to perceive something *despite* changes in the illuminant; this is worrisome insofar as constancy is sometimes revealed not *despite* changes in the illuminant, but *because of* changes in the illuminant. For example, if two patches are discriminable for me in some viewing condition, I might not know whether I have a case of colour constancy on my hands until I see how the visual appearances of the two change with respect to each other as the illumination changes uniformly on both of them.

scene more precisely, is called colour constancy ([Barnard, 1998], 1).[20]

Because reflectance perception accounts are formulated in terms of reflectance distributions, and because the latter (again, exceptions aside) have the right sort of invariance, these accounts are not vulnerable to falsification by empirical data in the way that invariantist accounts are (see §1).

That said, reflectance perception accounts are subject to their own sets of worries. In the rest of this section, I want to raise two kinds of worries for reflectance perception accounts. The first, which I take to be potentially solvable, is an undergeneration worry — a worry that reflectance perception accounts fail to count as instances of colour constancy cases that might reasonably be thought to be instances. The second, to which I don't see any solution, and which I consequently regard as much more serious, is that such accounts overgenerate instances of colour constancy — that they count as instances cases that should not be so treated.[21]

The undergeneration worry I have in mind is that the reflectance perception account of constancy is, on its face, inapplicable to instances of colour constancy involving transparent volumes, luminous sources, and films, insofar as the apparent colours (and colours) of these objects are not well accounted for in terms of spectral reflectance distributions. But it seems inappropriate to rule out these cases preemptively: some cases involving transparent volumes, luminous sources, and films seem to be (pending reasons to believe the contrary) *bona fide* examples of colour constancy.[22] A *prima facie* example involving a transparent volume can be generated by replacing the coffee cup in figure 1 with a transparent volume of beer (suppose the volume of beer is contained within a transparent glass). One example involving a luminous source — one that is standardly used in empirical investigations of colour constancy — involves adjacent CRT regions that fall under varying amounts of illumination by diffuse room lighting; another example is the firefly whose bio-luminescent body is partly illuminated and partly unilluminated by a flashlight. Finally, we can consider the case of a soap film that, like the coffee cup in figure 1, is partly in direct sunlight and partly in shadow. Each of these cases, I suggest, has a strong

---

[20]The difference between Barnard's and Palmer's formulations is less significant than it might seem, for the rest of Barnard's essay makes it clear that what he has in mind by "see[ing] the physical scene" is the extraction of surface spectral reflectance distributions.

[21] A third, more philosophically motivated worry about reflectance perception accounts arises from arguments to the effect that dispositional properties cannot be perceived (see, for example, [McGinn, 1996] and [Jackson, 1996]). Since reflectance distributions are dispositional properties, reflectance perception accounts threaten to entail the (presumably unwanted) consequence that colour constancy is (necessarily) never instantiated. On the other hand, for reasons I've discussed in [Cohen, 2003], I am not convinced by the worry that dispositional properties are imperceptible, so I don't take this objection to be decisive.

[22]I do not believe that such cases *must* be counted as instances of colour constancy. Rather, I am suggesting that they have a good *prima facie* claim to be so counted, and that this *prima facie* claim is something that should not be preempted stipulatively at the beginning of the day. This is intended to leave open the possibility that we might decide to revise our initial assessments at the end of the day.

prima facie claim to be counted an instance of colour constancy, by virtue of their similarity to the paradigm examples of colour constancy in figures 1–2. But, insofar as the colours of transparent volumes, luminous sources, and films cannot be understood in terms of surface spectral reflectance distributions, reflectance perception accounts of colour constancy preemptively rule out these cases.

The reason the undergeneration worry strikes me as not decisive is that there are plausible ways of extending the proposal to deal with the sorts of examples just mentioned. One way of carrying out this extension builds on the notion of *productances*, as defined by [Byrne and Hilbert, 2003]:

> Earlier we gave a standard definition of reflectance: the proportion of incident light the object is disposed to reflect at each wavelength in the visible spectrum. However, we could just as well have characterized reflectance slightly differently, in terms of the light that would *leave* the object, rather than the light that the object would *reflect*. ... [S]ay that the *productance* of a surface is its disposition to produce (reflect or emit or transmit) a specific proportion of incident light ([Byrne and Hilbert, 2003], 11).

In other words, if $r(\lambda)$ is the amount of light of wavelength λ reflected by an object O (surface, volume, whatever), $E(\lambda)$ is the amount of light of wavelength λ emitted by O, and $I(\lambda)$ is the amount of illumination of wavelength λ incident on O, then we can express the productance of O as the function that maps a wavelength λ to the ratio $(r(\lambda) + E(\lambda))/I(\lambda)$ (where λ ranges over the visible wavelengths). (So defined, the productance of a non-emitting surface will be equivalent to its spectral reflectance distribution; this legitimates the idea that productance amounts to a generalization of the original notion.) The thought, then, is that we could bring volumes, sources, and films into something in the spirit of a reflectance perception account of colour constancy by (mirroring the formulation from Palmer above) characterizing colour constancy as the ability to perceive the productance of objects despite changes in illumination.[23]

A concern that I think is ultimately more serious for reflectance perception accounts is that they overgenerate instances — that (unless they are restricted in such a way that they turn out to be versions of invariantism after all) they count as instances of colour constancy cases that they should not. To see this, I want to use figure 3 to describe a paradigm case of a failure of colour constancy. Figure 3 was derived by making a small change to the image in figure 1: a black border has been added to the region of the photograph that depicts the

---

[23][Jakab and McLaughlin, 2003] and [Mausfeld and Niederée, 2003] object to Byrne's and Hilbert's notion of productances on the grounds that the productance of an object relative to extremely low incident illumination (e.g., the aforementioned firefly at night) approaches infinity or is undefined. While this is a legitimate cause for concern, I don't want my objection against reflectance perception accounts to turn on this technicality — partly because I suspect that this problem might be finessed away with sufficient ingenuity (cf. [Byrne and Hilbert, 2003], 54), and partly because I think there is a more serious (i.e., not merely technical) problem with the account — viz., the overgeneration problem to be discussed below. Thus, I'll ignore the technical issue for the sake of argument.

Figure 3: Adding a well-placed border to figure 1 substantially weakens the impression that the two regions of the depicted coffee cup are alike in apparent colour.

sunlit region of the coffee cup. But this small change has the interesting effect of substantially weakening the impression that the two regions of the depicted coffee cup are alike in apparent colour. That is, figure 3 is unlike figure 1 in this crucial respect: most subjects have the reaction (regardless of whether they are asked to make appearance matches or surface matches; see §1) that the surface depicted in the left half of figure 3 (the shadowed region) is different in apparent colour from the surface depicted in the right half of the figure (the sunlit region). Presumably, then, this case should not come out as a case of colour constancy. But suppose we understand colour constancy as the perception of spectral reflectance distributions, as per a reflectance perception account. And suppose that (waiving the worries mentioned in note 21) we can and do perceive reflectance distributions. Then presumably we perceive reflectance distributions when we visually attend to the adjacent regions depicted in figure 3. (At least, there is no less reason for thinking that we perceive reflectance distributions when we attend to the adjacent regions depicted in figure 3 than there is for thinking that we perceive reflectance distributions when we attend to the corresponding regions depicted in figure 1.) If we do succeed in perceiving reflectance distributions in this case, then nothing has prevented us from perceiving reflectance distributions; *a fortiori* the change in illumination under which the coffee cup is viewed in the two halves of figure 3 has not prevented us from perceiving reflectance distributions when we look at the regions of the coffee cup. Thus, when visually attending to the coffee cup depicted in figure 3,

we perceive reflectance distributions despite changes in illumination. Consequently, on reflectance perception accounts of colour constancy, we experience an instance of colour constancy when we visually attend to the regions depicted in figure 3 — an outcome at odds with our initial assessment of the case as a paradigm failure of colour constancy.[24]

Now, a likely response to this objection is that it turns on an uncharitable reading of its target. Reflectance perception accounts, on this response, must not hold (despite the formulations in the sources quoted above) that colour constancy would involve *any* instance of the ability to perceive reflectance distributions despite changes in illumination. Rather, perhaps such accounts are intended to say something more like the following: a visual system exhibits colour constancy just in case it co-classifies samples that share a common spectral reflectance distribution, despite variations in the illumination. This modification does, in my view, manage to evade the difficulty pressed above: namely, it correctly classifies the failure of colour constancy involving figure 3 as a failure of colour constancy. However, the modified version of the proposal suffers from the problems brought out in our earlier discussion of invariantism (§1). For, as we have seen, subjects often fail to co-classify samples that share a common reflectance distribution when those samples are presented under different illuminations (namely, when they make appearance matches among those samples). Thus, proponents of the reflectance perception account appear to be faced with a dilemma: either they adopt the proposed modification, in which case they are susceptible to the difficulties that plague invariantism, or they reject the proposed modification, in which case they are stuck with an excessively permissive treatment of colour constancy. Neither option seems attractive.

3   Colour Constancy and Colour Ontology

We have seen two possible understandings of colour constancy, and noted that there are serious troubles for each of them — that there are reasons for worrying that neither of them provides an adequate characterization of the phenomenon of colour constancy. I now want to suggest that this has serious repercussions for certain prominent arguments about the nature of colour that turn on these

---

[24]Once again, I'm interested in the case where what is perceived is the coffee cup, not the photograph of the coffee cup (see note 3). Here, however, one might reasonably object that the border that distinguishes figure 3 from figure 1 is not part of the scene depicted, but only part of the representational vehicle. That is, one might worry that the crucial aspect of figure 3 on which the current objection depends cannot arise in any ecologically valid stimulus.

I'll answer this worry by describing what seems to me to be an ecologically valid case that has the structure I need. Suppose, then, that a subject views the partially lit coffee cup (the very coffee cup depicted in figures 1 and 3) through a clear glass panel upon which an opaque black quadrilateral has been painted. The subject views the coffee cup through the panel at an angle and distance so that the quadrilateral surrounds the entire region of the coffee cup that is under direct sunlight and visible. In short, the subject's view of the scene is much like your view of figure 3. The subject is perceiving the coffee cup, not the figure. But the opaque black border is also perceived and thereby has a perceptual effect on the subject that is equivalent to the effect that the black border in figure 3 has on you.

understandings of colour constancy.

To see why this is so, let us consider the appeals philosophers have made to colour constancy in the course of arguments to the conclusion that colours are objective, illumination-independent properties of some sort (e.g., surface spectral reflectance distributions). For example, here is Michael Tye: "The fact that objects appear to retain the same color through a wide variety of changes in illumination conditions (though certainly not all) strongly suggests that colors are illumination-independent properties of those objects" ([Tye, 2000], 147–148). Similarly, Hilbert begins with the claim that "The existence of color constancy should lead us to suspect the existence of some illumination-independent property of objects that is correlated with color," goes on to note that "The obvious candidate for such a property is . . . surface spectral reflectance" ([Hilbert, 1987], 65), and ends up concluding that colours are indeed identical to (classes of) surface spectral reflectance distributions. Finally, the thought is expressed nicely in the following passage from [Byrne and Hilbert, 2003]:

> We can narrow the field further by noting that the color vision of human beings and many other organisms exhibits approximate *color constancy* . . . ; for instance, tomatoes do not seem to change color when they are taken from a sunny vegetable patch into a kitchen illuminated with incandescent light. Assuming that our perceptions of color are often veridical, we therefore need a physical property of objects that is largely *illumination-independent* — a physical property that an object can retain through changes in illumination. . . . The property that initially suggests itself is *surface spectral reflectance*: the proportion of incident light the object is disposed to reflect at each wavelength in the visible spectrum ([Byrne and Hilbert, 2003], 9).

It seems that these philosophers are urging upon us something like the following chain of (non-deductive) reasoning (henceforth, the objectivist's inference):

(1) There is colour constancy.

(2) Reflectance distributions are illumination-independent.

(3) Therefore, colours are identical to reflectance distributions.

Our question, then, is whether either invariance accounts or reflectance accounts of colour constancy can sustain the objectivist's inference.

3.1 Invariance Accounts and Colour Ontology

Consider an invariantist reading of the objectivist's inference first.

On an invariantist account, premise (1) of the objectivist's inference amounts to the following, understood as an empirical claim:

(1′) The apparent colour of a surface is invariant across changes in illumination.

Of course, (1′) will not, by itself, give us any reason to believe the intended conclusion (3) unless we take (1′) to tell us something about *colour*, as opposed to *apparent colour*. The thought, then, must be that (1′) is a (defeasible) reason for believing this:

(1*) The colour of a surface is invariant across changes in illumination.

The objectivist's inference, on this reading, comes out as a pair of inferences to the best explanation. It moves first from the empirical premise (1′) to the effect that apparent colours are illumination-independent to the allegedly best explanatory conclusion (1*) to the effect that colours are illumination-independent; and from (1*) and the definitionally true premise (2) to the effect that reflectance distributions are illumination-independent, to the allegedly best explanatory conclusion (3), that colours are identical to reflectance distributions.

Understood in this way, the objectivist's inference invites a number of worries. First, there is a concern that, in the context of a debate about what the empirical results about apparent colour tell us about colour, the inference from (1′) to (1*) will strike many as too hasty (cf. note 8). Although I am sympathetic to this criticism, I'll put it aside in order to focus on what seem to me more fundamental problems with the inference that concern premise (1′) in particular.

One fundamental problem is that the results discussed in §1 provide reasons for suspecting that (1′), the key empirical premise of the argument, is false. Unless and until these suspicions can be allayed, we should reject the inference.

A second fundamental problem is that, even putting aside the variance/appearance match reactions discussed in §1 entirely, there are many uncontroversial cases — cases that even Hilbert and Tye would accept — where the alleged invariance cited in (1′) fails; perhaps the case involving the stimulus depicted in figure 3 is one of them (although nothing I say will turn on the assessment of this particular case). I take it that this much would conceded by all sides; certainly it is uncontroversial for colour scientists, who frequently take such cases to show the limitations on human colour constancy (thus, in an entirely typical passage, Peter Lennie writes that "Formal accounts of color constancy characterize mechanisms that perform better than human observers: Human color constancy is imperfect" ([Lennie, 1999], 245–246)).[25] Indeed, I take it that the parenthetical qualification in the quotation above from ([Tye, 2000], 147–148) is intended to concede that the claimed invariance fails in at least some cases.

But if the invariance fails in many cases, then we are justified in wondering why it is properly thought of as an invariance at all (rather than a variance). And given that Tye, Hilbert, and others are prepared to take the invariance of apparent colour across changes in illumination as evidence that (apparent colours, and therefore) colours are illumination-independent, why should we

---

[25]For more on the gap between invariance characterizations of colour constancy and the performance of human visual systems, see also [Mausfeld, 1998] and [Troost, 1998]. For a more extreme view, see [Foster, 2003], who argues that it is inappropriate to think of the human visual system as exhibiting colour constancy at all, insofar as that is taken to require recovering an invariant (i.e., illumination-independent) characterization of perceived surfaces.

not take the failures of invariance as showing that (apparent colours, and there-fore) colours are illumination-*dependent* properties of objects? Or, to put the point in terms of the inference form outlined above, our question is why, on an invariantist account of colour constancy, the empirical premise (1′) in the objec-tivist's inference is not refuted by the empirical evidence: why is the empirical claim that apparent colours are illumination-independent not disconfirmed by the (uncontroverted) cases where invariance fails?

Tye and Hilbert are aware of this worry, and their attempts to answer it are instructive. They propose to regard the violations of invariance as exceptional cases to the general rule (viz., the rule of invariance) on the grounds that these cases are cases of colour *mis*perception (cf. [Hilbert, 1987], 71–72, [Tye, 2000], 153ff).

Unfortunately, it is not clear to me that there are reasons for endorsing this view that are independent of the intended conclusion that colours are il-lumination-independent. Of course, if colours are assumed to be illumination-independent properties, then cases in which apparent colour is shown not to be illumination-independent are plausibly described as cases where the apparent colour fails to reveal the true colour of objects — as cases of colour mispercep-tion. But this result, of course, turns on assuming that colours are illumination-independent, and presumably this is not a supposition that we should make in our characterization of colour constancy if we hope to appeal to colour con-stancy to motivate the view that colours are illumination-independent. Nor, as far as I can see, is there any other reason in the offing for treating the uncontro-verted violations of invariance as ignorable exceptions. Thus, short of begging the ontological question that colour constancy is being enlisted to answer (via the objectivist's inference), it is not clear why the acknowledged failures of in-variance should not be regarded (*pace* Hilbert and Tye) as evidence for the illumination-dependence of (apparent colour, and therefore) colour.

3.2 Reflectance Accounts and Colour Ontology

If the invariantist understanding of colour constancy does not support the ob-jectivist's inference, we should ask whether that inference can be saved by re-liance on a reflectance perception account of colour constancy. On this reading, premise (1) in the inference is to be understood as this empirical claim:

(1″) We have the ability to perceive surface reflectance distributions (/produc-tances).

The question I now wish to consider is how, even granting the truth of premises (1″) and (2), the conclusion (3) is supposed to follow.[26]

---

[26]For the moment I'll put aside the further issue about how we should read 'we' in (1″). As I shall discuss in §4, there is evidence of colour constancy at work in the visual systems of infant humans and many non-human animals, *inter alia*. It is not clear whether the proponent of the current version of the objectivist's inference would want to apply the inference to these visual systems as well, although if not more would need to be said about why.

It is true that we may legitimately infer from (1″) and (2) that we have the ability to perceive something that is illumination-independent. But we have been given no reason, so far, for believing that that illumination-independent something we have the ability to perceive is colour. Consequently, we lack any reason, from what has been said so far, for endorsing (3). Of course, we would have reason for believing that the illumination-independent something we have the ability to perceive is colour if we helped ourselves to (3) as a premise; but, once again, availing ourselves of (3) as a premise would render the objectivist's inference question-begging. Consequently, availing ourselves of a reflectance perception understanding of colour constancy doesn't motivate the hoped-for identification of colours with reflectances.[27]

4   Colour Constancy as Counterfactual

So far I have argued that the usual ways of understanding colour constancy are unsatisfactory (§§1–2), and that they cannot serve the purposes to which philosophers have wanted to put them (§3). In this section I want to propose an alternative understanding of colour constancy, and then argue that it is superior to the construals considered above.

Our problem is to characterize the phenomenon involving our perception of the adjacent and simultaneously perceived regions of the coffee cup in figure 1 as well as the corresponding but successively perceived regions of figure 2. The discussion that has come so far suggests four desiderata that an acceptable account of colour constancy should meet. First, we want to say that colour constancy involves some sense in which the apparent colours of these regions are relevantly alike to subjects; this desideratum surely captures one intuition that we have about cases of colour constancy, and that comes out very clearly in invariance reactions/surface matches. Second, we need a way of articulating this last idea while (unlike the invariantist) avoiding the straight-out insistence that the regions are identical in apparent colour, since, as shown by the variance reactions/appearance match data, there is a good sense in which they are not. Third, the account we arrive at ought to be extensionally adequate: it ought not be vulnerable to the problems of undergeneration and overgeneration that made trouble for the reflectance perception account in §2. And fourth, we should provide an explanation of why the different instructions used to elicit appearance matches and surface matches lead subjects to provide the responses they do, and how the two sorts of reactions are related to one another.

I propose to address these desiderata by claiming that, in cases of colour constancy, one of the responses of visual systems amounts to an answer to a question about the counterfactual properties of the regions under comparison. Namely, these visual systems answer this question: would region R_1 (presented

---

[27]The problem I'm urging here is *not* that (1″) and (2) fail to provide *deductive* reasons for endorsing (3); i.e., I'm not merely complaining about the reliance on non-deductive forms of argument (I take it that that sort of complaint is itself unmotivated). It is, rather, that (1″) and (2) fail to provide *any* rational grounds for endorsing (3).

under illumination I_1) share an apparent colour with region R_2 (presented under illumination I_2) if, contrary to fact, both regions were presented under the same illumination — namely, both under I_1 or both under I_2?[28]

Thus, on this proposal, which I'll call *counterfactualism* about colour constancy, the respect in which the two regions are alike — what is constant in cases of colour constancy — is not their occurrent apparent colour, but their counterfactual apparent colour. The visual system's responsiveness to this counterfactual dimension of comparison drives one of our reactions to cases of colour constancy (namely, the invariance/surface match reaction). On the other hand, I claim that our visual systems are also responsive to the distinct dimension of occurrent apparent colour, and so can generate answers to this question: does region R_1 (presented under illumination I_1) currently share an apparent colour with region R_2 (presented under illumination I_2)? And, I suggest, it is our visual systems' responsiveness to this distinct dimension of comparison that drives the other of our reactions to cases of colour constancy (namely, the variance/appearance match reaction). Putting all this together, counterfactualism understands colour constancy as the capacity of the visual system to discern similarity in counterfactual apparent colour across differences in occurrent apparent colour.

Below I'll defend counterfactualism against objections. Before I come to this, however, I want to argue that the view respects the desiderata cataloged above.

4.1   Advantages of Counterfactualism

First, of course, counterfactualism offers a direct account of the sense in which the two regions of interest in a case of simultaneous or successive colour constancy — regions that are revealed to be alike under surface match conditions — are alike in respect of colour. It does not say that such regions are alike in that they share an apparent colour. Rather, it says that the two regions are alike in that they *would* share an apparent colour if, contrary to fact, both regions were presented under the same illumination (namely, under I_1 or under I_2). The apparent colour that the patches would manifest were they presented under I_1 or I_2 is a perfectly good dimension of comparison, and the present proposal has it that this is one relevant dimension along which visual systems represent pairs of patches such as those we've been discussing as being similar.

Second, and unlike invariantism, counterfactualism explains this likeness while simultaneously respecting the facts about variance/appearance reactions. This is because, unlike the invariantist, the counterfactualist allows that the two regions of interest in a case of colour constancy can (i) appear relevantly similar in respect of colour (by sharing a counterfactual apparent colour), even though

---

[28]It is significant that this last question is asked about apparent colours under I_1 and I_2. In particular, it won't do to ask whether the patches would be alike in apparent colour were they presented under every common illuminant (that constraint is too strong) or whether they would be alike in apparent colour were they presented under some common illuminant (that constraint is too weak).

(ii) they differ in apparent colour.[29] Given that this combination of verdicts is observed, it is a virtue of counterfactualism that it allows for the combination, and it is a vice of invariantism that it does not.

Third, counterfactualism avoids the problems of undergeneration discussed in §2 in a particularly simple way. For, unlike reflectance perception accounts, counterfactualism allows that there could be cases of colour constancy involving transparent volumes, luminous sources, and films: so long as these have apparent colours under various conditions of illumination, there will be a well-formed question whether volume (etc.) V_1 (presented under illumination I_1) and volume (etc.) V_2 (presented under illumination I_2) would share an apparent colour if, possibly counter to fact, they were both presented under I_1 or I_2.

In addition, the counterfactualist avoids the problems of overgeneration discussed in §2 because she will not count the perception of the coffee cup depicted in figure 3 as a case of colour constancy. On her view, the case is not an instance of colour constancy because one of the two responses that constitute colour constancy is absent. In particular, subjects' visual systems do *not* represent that the two regions depicted here would match in apparent colour if, contrary to fact, they were presented under the same illumination — rather, they represent that the two regions would *not* match in colour appearance if they were presented under the same illumination. As it happens, this conclusion about apparent colour is erroneous, and this explains why the case is naturally counted a *failure* of colour constancy. (On the other hand, counterfactualism invites questions about other sorts of undergeneration and overgeneration that deserve consideration on their own; I'll discuss these in §4.3.)

Fourth, counterfactualism provides the resources for explaining why the two particular experimental instructions used in asymmetric colour matching experiments generate the particular patterns of response that they do — why instructions to "adjust the test patch to match its hue and saturation to those of the standard patch" lead subjects to make appearance matches, while instructions to "adjust the test patch to look as if it were 'cut from the same piece of paper' as the standard, i.e., to match its surface color" lead subjects to make surface matches, and even provides an account of the relation between these two patterns of response.

Consider appearance matches first. The counterfactualist will hold that the surfaces of interest in cases of colour constancy — such as that depicted in figure 1 — differ in apparent colour. Assuming, with current orthodoxy in colour

---

[29] Indeed, it is reasonable to think that a visual system cannot represent (i) without also representing (ii): it is hard to see how visual systems could reach a verdict about apparent colours in counterfactual circumstances except on the basis of apparent colours in actual circumstances. Compare:

> 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. That is, perfect constancy could still obtain if the viewer, by a perceptual computation, were able to see the paper as an object of the same surface color under illumination perceived to be greener than the direct sunlight ([Arend and Reeves, 1986], 1749).

science, that apparent colour supervenes on hue, saturation and lightness, and since the samples are alike in lightness in experimental tests of surface colour constancy, it follows that the two patches fail to match in hue or saturation.[30] Consequently, when subjects are given the instruction to "adjust the test patch to match its hue and saturation to those of the standard patch," they do just that — they adjust the hue and saturation of the test patch until there is no longer a difference in hue or saturation between it and the standard patch, hence no longer a difference in apparent colour between it and the standard patch. That is, they make appearance matches.

In contrast, suppose the subject is given the instruction to "adjust the test patch to look as if it were 'cut from the same piece of paper' as the standard, i.e., to match its surface color." As before, the counterfactualist will hold that there is a difference in apparent colour between the test and standard patches. But unlike the prior situation, it is not at all clear that the way in which the subject can satisfy the new experimental instruction is to change the hue and saturation of the test patch until that difference goes away (i.e., to make an appearance match). After all, if two patches cut from a uniform piece of paper really were presented under different illumination, they would *not* be a perceptual match (in hue and saturation, hence in apparent colour) for the subject. Rather, they would differ in apparent colour but meet the further condition that *if, counterfactually*, the test patch and the standard patch were presented under the same illumination, *then* the two would match in apparent colour. Thus, the way for the subject to satisfy the experimental instruction in surface match conditions is not to make the test and standard patch match in apparent colour, but to adjust the test until her visual system represents that the two *would* match under the counterfactual condition in which they are presented under uniform illumination.

A final consideration in support of counterfactualism is that, unlike invariantism, it provides a plausible understanding of the relationship between invariance/surface match reactions and variance/appearance match reactions. In note 29 I pointed out that, if counterfactualism is correct, then invariance/surface match reactions are naturally understood as results computed by the visual system from the evidence of non-matching apparent colours. If this is right, then invariance/surface match reactions are the results of computations that have as inputs the outputs of variance/appearance match reactions. Consequently, counterfactualism predicts that appearance matches should be less difficult and faster than surface matches. I take the confirmation of these

---

[30]In recent years, a number of colour scientists have posed an important challenge to this orthodoxy by proposing that models of apparent colour need to be expanded to include information about the illuminant; see, for example, [Ekroll et al., 2002], [Ekroll et al., 2004], [Mausfeld, 2003], and, for earlier precedents, [Katz, 1911] and [Bühler, 1922]. Needless to say, this view is consonant with the aims of the present paper; indeed, the proposal I am making can be understood as one (admittedly schematic) way of making out part of the suggested view. However, because presupposing the suggested (unorthodox) view of colour appearance would beg precisely the questions I am attempting to resolve, it serves my dialectical and expository purposes not to make this presupposition. Thus, in the interests of ecumenicism, I'll proceed as if the orthodoxy were correct.

predictions ([Arend and Reeves, 1986], 1747–1748) as a further piece of support for the view.[31]

4.2 Psychological Reality

I have argued that counterfactualism offers an understanding of colour constancy that is theoretically and empirically well-motivated, and that it succeeds where competing accounts fail. However, the attribution of counterfactual properties — on which the counterfactualist's treatment of colour constancy rests — is presumably a fairly sophisticated intellectual feat. And this point leads to a number of interrelated concerns about the psychological plausibility of the counterfactualist treatment of colour constancy. The general worry here is that, in construing colour constancy as involving representing counterfactual properties of objects, the view over-intellectualizes a phenomenon that is found in such apparently cognitively unsophisticated creatures as goldfish, honeybees, (and several other non-human animals; see the review in [Neumeyer, 1998]) and very young human infants (somewhere between 9 and 20 weeks [Dannemiller and Hanko, 1987], [Dannemiller, 1989]).

The first point to make in this connection is that the sophistication of the representations and computational processing posited by counterfactualism is, by itself, unobjectionable. For, if the history of cognitive/perceptual science has taught us anything, it is that the cognitive and perceptual behavior of even the simplest creatures requires extremely sophisticated representational structures and processing defined over those structures. Moreover, we have learned that much sophisticated processing goes on beneath the level of introspective awareness; consequently, the lack of first-person introspective evidence for the complex processing and representation of counterfactual properties required by counterfactualism is no objection to that account.

Recall that the counterfactualist description of colour constancy is given in terms of two separate components corresponding to two distinct visual reactions: the visual system represents that relevant regions are alike in counterfactual apparent colour (a verdict reached by one component process) despite being unalike in occurrent apparent colour (which verdict is reached by another component process). On this story, the representation of counterfactual properties occurs in one of the components — namely, the component that explains our invariance/surface match reaction. I suggest that we should think of both of these components as neo-Helmholtzian subpersonal mechanisms for generating conclusions about the world on the basis of current visual input — mechanisms whose operations are not accessible to conscious introspection.[32] On the

---

[31]To be fair, the reports about the relative ease of appearance matches in [Arend and Reeves, 1986] and [Bäuml, 1999] are fairly anecdotal. I am not aware of any systematic, quantitative investigations of this issue.

[32]Moreover, these mechanisms exhibit most of the other classic features of mental modules, in the sense of [Fodor, 1983]: they are domain-specific, their operation is plausibly mandatory, fast, informationally encapsulated, shallow in outputs, susceptible to selective impairment, etc. While I cannot defend here for reasons of space these individual claims about colour constancy and the markers of modularity, I should mention that evidence for the selective impairment of

other hand I propose that the outputs of those mechanisms *are* accessible to conscious introspection: this is why subjects (and not just their subpersonal computational mechanisms can produce (and recognize in themselves) both invariance/surface match and variance/appearance match reactions.[33]

Thinking of colour constancy in this way also explains how our performance on colour constancy task can be responsive to instructional differences (e.g., the difference between an instruction to make surface matches and an instruction to make appearance matches), despite involving subpersonal (and, therefore, informationally encapsulated) mechanisms. The components themselves cited by the counterfactualist, insofar as they are informationally encapsulated, should be insensitive to instructional effects. But given that (according to counterfactualism) an organism has at its disposal both subpersonal mechanisms, its decision to engage one of them or the other, or to report on the outputs of one of them or the other, might very well occur at the personal level, and so be susceptible to instructional effects. By way of analogy, consider that subjects can choose between alternative 3-dimensional spatial interpretations of the 2-dimensional Necker cube configuration by choosing to treat one or another 2-dimensional region as the facing surface in 3-dimensional space. Indeed, this choice can even be influenced by instructions (e.g., the instruction to treat the region with faint green diagonal lines as the facing surface). But once this choice is made, subjects' subpersonal mechanisms for form perception get started, and the spatial interpretation of other regions become obligatory (unless and until the initial choice is reversed). As I hope this comparison makes clear, there is no clash between my claim that a subpersonal attribution of counterfactual properties is at work in implementing the invariance/surface match response and my claim that our reactions to cases like that depicted in figures 1–2 are subject to instructional influence.[34]

Perhaps it will nonetheless strike some readers as objectionable to claim that any subpersonal mechanism — e.g., in the visual systems of non-human animals and human infants — can represent the kinds of counterfactual properties that I've invoked to explain part of our responses to cases of colour constancy. Of course, the counterfactualist is not claiming that colour constancy requires (hence, that goldfish and human infant visual systems have) the capacity for the sort of wide-ranging, domain-neutral, explicitly articulated counterfactual judgments that normal adult human beings have. Rather, she is claiming that colour

---

colour constancy (which I take to be the most controversial of the criteria listed) is presented in [Rüttiger et al., 1999].

[33]The strategy of explaining personal-level behavior in terms of the outputs of subpersonal-level computational mechanisms is ubiquitous in cognitive science (which is not to say that it is without problems), and so not a special problem for the counterfactualist. Thus, it is whole persons who produce the verdicts (verbally, by button-presses, and so on) that various strings of words are grammatical or not, or that a figure is distinct from a ground in a geometric configuration, or that one line is longer than another (etc.) in an experimental setup; but the standard story is that these verdicts are the outward, personal-level expressions of outputs from subpersonal mechanisms that can and should be studied on their own. If this strategy serves to explain linguistic competence or form perception (*inter alia*), it should suffice for explaining colour constancy as well.

[34]Thanks to an anonymous referee for prompting me to clarify my views on this point.

constancy requires (hence that goldfish and human infant visual systems have) a capacity for a subdoxastic, cognitively impenetrable, and domain-specific sort of counterfactual attribution.[35] Exactly what form this counterfactual attribution takes within the visual system, and exactly how the visual system computes such counterfactual verdicts from occurrent colour appearances, are hotly debated matters within the field of computational colour constancy that go beyond the scope of this paper (see [Hurlbert, 1998] for a systematic overview of this large and active area of research). What matters for our purposes is that, given that the counterfactualist's understanding of the capacity for the representation of counterfactual properties is restricted in the ways discussed, there is no longer any obvious reason for objecting to the attribution of that capacity to subpersonal mechanisms such as we might find in the visual systems of goldfish and human infants.

For the reasons canvassed in this section, I do not believe that considerations about psychological plausibility in general, and the developmental or comparative data in particular, tell against the counterfactualist account of colour constancy. On the contrary, I am inclined to think that the view is quite psychologically plausible, and may turn out to be useful in understanding the ontogeny and phylogeny of colour constancy in a wide range of visual systems other than our own sort.

4.3 Undergeneration and Overgeneration

As noted above, counterfactualism invites its own under- and overgeneration concerns.

One such undergeneration concern has already been addressed in §4.2 — viz., the concern (related to the undergeneration concerns raised for reflectance perception accounts in §2) that at least some visual systems that seem to exemplify colour constancy might seem incapable of representing the complicated counterfactual properties required by the account. As I argued in §4.2, there are reasons for believing that the sorts of counterfactual property ascriptions required by counterfactualism are within the reach of the visual systems that exhibit colour constancy, and consequently for believing that these sorts of undergeneration concerns can be met.

Another class of undergeneration worry for counterfactualist is connected with the finkish dispositions mentioned in note 18. Suppose a visual system is

---

[35][Harris et al., 1996] (see also [Harris, 2000]) give evidence for the claim that that 3-year-old human beings do in fact have a capacity for wide-ranging, domain-neutral, counterfactual judgments explicitly expressed in natural language (these results have been challenged by [Riggs et al., 1998] and [Riggs and Peterson, 2000], but see [German and Nichols, 2002] for a convincing reply). Of course, this result doesn't, by itself, support the contention about the capacities for counterfactual attribution in 9–20 week-old human infants or the non-human organisms in whom colour constancy is present. On the other hand, if this result gives reason to believe that 3-year-old human beings possess a capacity for counterfactual judgment that is domain-general, and that depends on understanding syntactically complex expressions in natural language, then I hope it will soften resistance to the attribution of capacities for much more domain-limited and language-independent representation of counterfactual properties to the visual systems of even younger human infants and non-linguistic creatures.

presented with a pair of regions under a pair of illuminations, as in a standard asymmetric colour matching setup: a region R_1 of previously unexposed photosensitive paper illuminated by a dim red safelight I_1 (viz. the sort of light used in photographic darkrooms to provide ambient illumination without causing photosensitive paper to become exposed), and a region R_2 of standard white office paper illuminated by direct sunlight I_2. R_1 and R_2 are, of course, presented under distinct illuminants. But here is a true counterfactual conditional: if R_1 and R_2 were presented under a common illuminant I_2, they would fail to be a perceptual match, insofar as R_1 would appear much darker than R_2 were it presented under I_2. The reason this case can be used to mount an undergeneration worry for counterfactualism is that it seems as if the visual system will represent the pair of R_1 under I_1 and R_2 under I_2 as similar in just the way we expect of an instance of colour constancy. But given the counterfactuals we've uncovered about the pair, it looks doubtful that any counterfactual analysis — viz., any account that requires that the pair would share an apparent colour were they presented under I_2 — can count the case as an instance of colour constancy.

The reason this objection does not derail counterfactualism is that that view, as I have presented it, does not make any such requirement. To count as an instance of colour constancy for a visual system V, counterfactualism does not require that R_1 and R_2 *would* be a perceptual match (for V) were the two (counterfactually) both presented under I_2; what it requires is that visual system V *represent* that the pair of regions would be a perceptual match (for V) were they (counterfactually) both presented under I_2. As far as I can see, there is no reason not to believe that this requirement would not be met in the case we are discussing; indeed, the satisfaction of this requirement looks like the natural way of describing the case.[36] Consequently, it seems that the case will be counted by counterfactualism as an instance of colour constancy (as desired).

Of course, if it is, then this means that visual systems for which the pair is an instance are representing the world erroneously: they represent the pair of surfaces as standing in a counterfactual relation that, in fact, they fail to stand in. But this, too, strikes me as the right thing to say about the case, and so

---

[36]Objection: Suppose that at t_1 I believe/represent that R_1 and R_2 would be a perceptual match for that my visual system were the two (counterfactually) both presented under I_2. But by a later time t_2, I have learned about how photographic paper works, and become adept at recognizing this sort of paper (and safelight illumination) when I encounter it. Thus, at t_2, when presented with the same pair, I will *not* believe/represent that R_1 and R_2 would be a perceptual match for that my visual system were the two (counterfactually) both presented under I_2. Then, on the present account, the pair counts as an instance of colour constancy at t_1 but not at t_2, merely as a function of what I know about photographic paper. But surely whether a pair is an instance of colour constancy (for me) should be independent of my detailed knowledge about photographic equipment (and my other beliefs, more generally speaking).

Reply: Whether the pair counts as an instance, on the counterfactualist account, turns on the counterfactual properties *visual systems* represent about that pair, not what *subjects* represent about that pair. Learning about photographic paper and the like affect the latter, but not the former; by way of analogy, measuring the lines in a Müller-Lyer illusion may change what I represent about their relative lengths, but it fails to alter what my visual system represents about their relative lengths (cf. [Pylyshyn, 1999]). For this reason, counterfactualism does not make the instancehood of the pair dependent on my general beliefs.

not a defect of counterfactualism: after all, the finkish features at the heart of the case mean that things will change in ways that are unanticipated (hence conducive to error) when we alter the illumination. Thus, counterfactualism will not only correctly classify the case as an instance of colour constancy, but will also point toward an explanation of exactly what it is about the case that visual systems misrepresent.

The same ingredients of the undergeneration worry just discussed can also be used to raise an overgeneration worry. Once again, suppose a visual system is presented with a pair of regions under a pair of illuminations, as in a standard asymmetric colour matching setup: this time, we have region R_1 of previously unexposed photosensitive paper illuminated by a dim red safelight I_1 (as before), and region R_3 of black construction paper illuminated uniformly by direct sunlight I_3. As before, R_1 and R_3 are presented under distinct illuminants. Once again, because of the finkish nature of R_1, the counterfactuals are otherwise than a visual system would typically represent. Namely, if R_1 and R_3 were presented under a common illuminant I_3, they *would be* a perceptual match, insofar as R_1 would appear much darker were it presented under I_3 than it does when (as it is in fact) presented under I_1.

Here I think we should expect that the case should not count as an instance, insofar as the reactions a visual system has to the pair of R_1 under I_1 and R_3 under I_3 seem unlike the reactions a visual system has to pairs in canonical instances (e.g., to the pairs in figures 1–2). The overgeneration concern will be, then, that (since R_1 and R_3 *would* be a perceptual match were they presented under I_3), any counterfactual-based account will wrongly classify the case as an instance of colour constancy.

But this concern can be answered. For, as I pointed out in connection with the undergeneration case, the truth or falsity of the counterfactuals about matching is, by itself, insufficient to determine whether some visual system will exhibit an invariance/surface match reaction, hence insufficient (even given that the visual system exhibits a variance/appearance match reaction) to compel a verdict about whether the case is or is not an instance of colour constancy according counterfactualism. Rather, what matters is what visual systems represent about the counterfactual properties of the surfaces.[37] And it seems that counterfactualism gets the right answer here: a normal visual system will (although a sophisticated believer might not) represent that the pair fails to satisfy the counterfactual relation. Once again, this means that visual systems misrepresent the world; but, also once again, this seems appropriate given the finkish

---

[37]It should be clear that the response I am offering here is more general than the specific cases at hand; in particular, it works generally against the cases brought to light by [Shope, 1978] that are sometimes thought to be damning to counterfactual theories of almost anything. The present proposal is not vulnerable to such cases because (unlike the counterfactual accounts that Shope takes as his target) it only adverts to counterfactuals that fall inside the scope of what is represented by visual systems. Since visual systems can and do persist in representing such counterfactuals when (for Shope-type/finkish reasons) they are false, or persist in failing to represent these counterfactuals when (for Shope-type/finkish reasons) they are true, the verdicts of counterfactualism about individual cases come apart from the counterfactuals that hold of the cases.

details of the case.

5   Ontological Implications

As we have seen, colour constancy has been used to argue for certain views about the ontology of colour properties — in particular, to argue for the view that colours are identical to spectral reflectance distributions (§3). I have suggested that these attempts to extract philosophical conclusions from colour constancy are unsuccessful. However, it is worth returning, at this point, to ask what implications the counterfactualist account of colour constancy advocated in §4 has for colour ontology.

As discussed in §4, counterfactualism is built to accommodate the compatibility of these two claims:

- region R_1 presented under illumination I_1 differs in apparent colour from surface region R_2 presented under illumination I_2,

- region R_1 presented under illumination I_1 shares an apparent colour with surface region R_2 presented under illumination I_1.

And once again, it is to the credit of counterfactualism that it respects both of these claims, since the empirical results suggest strongly that, in ordinary cases of colour constancy, both are true. Of course, if these claims are both true, it follows that the apparent colour of a surface region is illumination-*dependent* rather than illumination-*independent*.

However, the question that has made colour constancy worth caring about for most philosophers is not whether *apparent colours* are illumination-independent, but whether *colours* are illumination-independent. Like proponents of other accounts (see §3), I think that facts about apparent colour bear on the question of interest about colour; but I disagree with those other authors about the specific conclusions that are warranted. Indeed, I think our reflections on apparent colour give us (defeasible) reasons for thinking that colours are illumination-dependent. There are two ways in which we can motivate this conclusion.

First, recall that, for several authors, the (alleged) illumination-independence of apparent colour is itself reason for thinking that colours are illumination-independent as well (by way of some sort of inference to the best explanation). But if, as we have seen, a fuller description of the data suggests that apparent colours are illumination-dependent, then this gives an equally good (or bad) inference to the best explanation argument to the conclusion that colours are illumination-dependent.

A second argument from the illumination-dependence of apparent colour to the illumination-dependence of colour begins by supposing, for *reductio*, that colours are illumination-independent — that the colour of surface region R is stable across changes in illumination. If the apparent colour of R varies with the illumination, then our supposition (together with standard assumptions) entails that at most one of the distinct apparent colours that R has across a range of

illuminations veridically represents the colour that R has. Thus, the defender of the thesis that colours are illumination-independent owes us a principled reason for singling out (at most) one of the various apparent colours that R has as the unique apparent colour that veridically represents R's colour. As I have argued elsewhere ([Cohen, 2000], [Cohen, 2004]) it seems to me unlikely that such a principled reason is forthcoming. If it is not, then this gives us reason for rejecting the supposition that colours are illumination-independent.[38]

To be sure, these two arguments for the illumination-dependence of colour are non-deductive, and certainly open to challenge. But whether one accepts these arguments or not, it is important to see that the phenomenon of colour constancy is at least consistent with (and arguably supportive of) an illumination-dependent conception of colour. After all, many writers have claimed that colour constancy is incompatible with illumination-dependent conceptions of colour properties, or that by itself it gives strong support to the view that colours are illumination-independent (e.g., [Tye, 2000], 147–148, [Hilbert, 1987], chapter 5). In light of what we have said, these latter conclusions seem to be incorrect.

Counterfactualism about colour constancy, then, not only (i) corrects inadequate understandings of the phenomenon, but (ii) undercuts arguments from that phenomenon to the conclusion that colours are illumination-independent, and (iii) is fully compatible with the view that colours are illumination-dependent.

6   Conclusion

I've argued that the usual understandings of colour constancy in the philosophical and empirical literature are unsatisfactory, and that the counterfactualist account defended here can provide the needed correction. This result matters for at least two reasons. First, colour constancy is a fascinating phenomenon in its own right, and we need a way of characterizing the phenomenon that is adequate to the data. Second, colour constancy has served as a crucial line of empirical support for objectivist views about colour ontology (such as the prominent view that colours are identical to spectral reflectance distributions),

---

[38]Objection: "We simply do not say of objects that they change in colour under different lighting conditions.... And whatever colours are, it would be a violation of our ordinary concept of colour to say that an object changes colour when a passing object casts it in shadow" ([Thompson, 2006], 89–90).

Reply: Whether an object such as the coffee cup of figure 1 exemplifies one colour or two cannot be decided by "common sense" or "our ordinary concept of colour" as accessed from armchair reflection alone. As we have seen, our best qualitative and quantitative operational tests for individuating colours (i.e., perceptual matching tests) give conflicting verdicts about cases of interest: sometimes subjects react in a way that is consistent with the coffee cup's exemplifying just one colour, and other times in a way consistent with its exemplifying two. It seems to me inappropriate to react to this situation by insisting that one of these reactions must be rejected, or that only one of them is worthy of being attached to the name 'colour'. Rather, we should develop a theoretical account that (like counterfactualism) recognizes *both* of the individuative standards we see at work, explains how they are related, and says why each functions in certain cases but not others (consistent with the observed data).

and I've argued that the apparent force of this support evaporates once a more adequate understanding of the phenomenon is in place.

Colour constancy has many lessons to teach us about the nature of colour and colour vision. Like many other things, we stand to learn much from this phenomenon by looking at it in a different light.[39]

References

[Arend and Reeves, 1986] Arend, L. and Reeves, A. (1986). Simultaneous color constancy. *Journal of the Optical Society of America A*, 3(10):1743–1751.

[Arend et al., 1991] Arend, L., Reeves, A., Schirillo, J., and Goldstein, R. (1991). Simultaneous color constancy: patterns with diverse Munsell values. *Journal of the Optical Society of America A*, 8:661–672.

[Barnard, 1998] Barnard, K. (1998). Modeling scene illumination colour for computer vision and image reproduction: A survey of computational approaches. Ms., Simon Fraser University.

[Bäuml, 1999] Bäuml, K.-H. (1999). Simultaneous colour constancy: how surface color perception varies with the illuminant. *Vision Research*, 39(8):1531–1550.

[Beck, 1972] Beck, J., editor (1972). *Surface Color Perception*. Cornell University Press, Ithaca.

[Blackwell and Buchsbaum, 1988] Blackwell, K. T. and Buchsbaum, G. (1988). Quantitative studies in color constancy. *Journal of the Optical Society of America A*, 5:1772–1780.

[Brainard et al., 2003] Brainard, D. H., Kraft, J. M., and Longere, P. (2003). Color constancy: Developing empirical tests of computational models. In Mausfeld, R. and Heyer, D., editors, *Colour Perception: Mind and the Physical World*, pages 307–328. Oxford University Press, New York.

[Bühler, 1922] Bühler, K. (1922). Die Erscheinungsweisen der Farben. In Bühler, K., editor, *Handbuch der Psychologie, I.Teil, Die Struktur der Wahrnehmungen*, pages 1–201. Fischer, Jena.

[Byrne and Hilbert, 1997a] Byrne, A. and Hilbert, D. R. (1997a). Colors and reflectances. In Byrne, A. and Hilbert, D. R., editors, *Readings on Color, Volume 1: The Philosophy of Color*, pages 263–288. MIT Press, Cambridge, Massachusetts.

[Byrne and Hilbert, 1997b] Byrne, A. and Hilbert, D. R. (1997b). *Readings on Color, Volume 2: The Science of Color*. MIT Press, Cambridge, Massachusetts.

[Byrne and Hilbert, 2003] Byrne, A. and Hilbert, D. R. (2003). Color realism and color science. *Behavioral and Brain Sciences*, 26(1):3–64.

[Cohen, 2000] Cohen, J. (2000). *Color Properties and Color Perception: A Functionalist Account*. PhD thesis, Rutgers University, New Brunswick, New Jersey.

[Cohen, 2003] Cohen, J. (2003). Color: A functionalist proposal. *Philosophical Studies*, 112(3):1–42.

[Cohen, 2004] Cohen, J. (2004). Color properties and color ascriptions: A relationalist manifesto. *The Philosophical Review*, 113(4):451–506.

[Cohen and Meskin, 2004] Cohen, J. and Meskin, A. (2004). On the epistemic value of photographs. *Journal of Aesthetics and Art Criticism*, 62(2):197–210.

[Cornelissen and Brenner, 1995] Cornelissen, F. W. and Brenner, E. (1995). Simultaneous colour constancy revisited: an analysis of viewing strategies. *Vision Research*, 35:2431–2448.

[Craven and Foster, 1992] Craven, B. J. and Foster, D. H. (1992). An operational approach to colour constancy. *Vision Research*, 32(7):1359–1366.

[Dannemiller, 1989] Dannemiller, J. L. (1989). A test of color constancy in 9- and 20-week-old human infants following simulated illuminant changes. *Developmental Psychology*, 25:171–184.

[Dannemiller and Hanko, 1987] Dannemiller, J. L. and Hanko, S. A. (1987). A test of color constancy in 4-month-old human infants. *Journal of Experimental Child Psychology*, 44:255–267.

[Delahunt, 2001] Delahunt, P. B. (2001). *An Evaluation of Color Constancy Across Illumination and Mutual Reflection Changes*. PhD thesis, University of California, Santa Barbara, Santa Barbara.

[Delahunt and Brainard, 2004] Delahunt, P. B. and Brainard, D. H. (2004). Does human color constancy incorporate the statistical regularity of natural daylight? *Journal of Vision*, 4:57–81.

[Ekroll et al., 2004] Ekroll, V., Faul, F., and Niederée, R. (2004). The peculiar nature of simultaneous colour contrast in uniform surrounds. *Vision Research*, 44(15):1765–1786.

[Ekroll et al., 2002] Ekroll, V., Faul, F., Niederée, R., and Richter, E. (2002). The natural center of chromaticity space is not always achromatic: A new look at colour induction. *Proceedings of the National Academy of Sciences*, 99(20):13352–13356.

[Evans, 1948] Evans, R. M. (1948). *An Introduction to Color*. Wiley, New York.

[Fodor, 1983] Fodor, J. A. (1983). *The Modularity of Mind*. MIT Press, Cambridge, Massachusetts.

[Foster, 2003] Foster, D. H. (2003). Does colour constancy exist? *Trends in Cognitive Science*, 7(10):439–443.

[German and Nichols, 2002] German, T. P. and Nichols, S. (2002). Children's counterfactual inferences about long and short causal chains. *Developmental Science*. In press.

[Goldstein, 1999] Goldstein, E. B. (1999). *Sensation & Perception (5th Edition)*. Brooks/Cole Publishing, Pacific Grove, California.

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

[Harris, 2000] Harris, P. L. (2000). *The Work of the Imagination*. Blackwell, Oxford.

[Harris et al., 1996] Harris, P. L., German, T. P., and Mills, P. E. (1996). Children's use of counterfactual thinking in causal reasoning. *Cognition*, 61:233–259.

[Hilbert, 1987] Hilbert, D. R. (1987). *Color and Color Perception: A Study in Anthropocentric Realism*. CSLI, Stanford.

[Hurlbert, 1998] Hurlbert, A. C. (1998). Computational models of color constancy. In Walsh, V. and Kulikowski, J., editors, *Perceptual Constancy: Why Things Look as They Do*, pages 283–322. Cambridge University Press, Cambridge.

[Hurvich, 1981] Hurvich, L. M. (1981). *Color Vision*. Sinauer Associates, Sunderland, Massachusetts.

[Jackson, 1996] Jackson, F. (1996). The primary quality view of color. *Philosophical Perspectives*, 10:199–219.

[Jakab and McLaughlin, 2003] Jakab, Z. and McLaughlin, B. (2003). Why not color physicalism without color absolutism? *Behavioral and Brain Sciences*, 26(1):34–35.

[Jameson and Hurvich, 1989] Jameson, D. and Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology*, 40:1–22. Reprinted in [Byrne and Hilbert, 1997b], 177–198.

[Katz, 1911] Katz, D. (1911). *Die Erscheinungsweisen der Farben und ihre Beeinflussung durch die Individuele Erfahrung*. Barth, Leipzig.

[Lennie, 1999] Lennie, P. (1999). Color coding in the cortex. In Gegenfurtner, K. R. and Sharpe, L. T., editors, *Color Vision: From Genes to Perception*, pages 235–247. Cambridge University Press, Cambridge.

[Maloney, 1986] Maloney, L. T. (1986). Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *Journal of the Optical Society of America*, A 3:1673–1683.

[Martin, 1994] Martin, C. B. (1994). Dispositions and conditionals. *Philosophical Quarterly*, 44(174):1–8.

[Mausfeld, 1998] Mausfeld, R. (1998). Color perception: From Grassman codes to a dual code for object and illumination colors. In Backhaus, W. G. K., Kliegl, R., and Werner, J. S., editors, *Color Vision: Perspectives from Different Disciplines*, pages 219–250. Walter de Gruyter, Berlin.

[Mausfeld, 2003] Mausfeld, R. (2003). The dual coding of colour: 'surface colour' and 'illumination colour' as constituents of the representational format of perceptual primitives. In Mausfeld, R. and Heyer, D., editors, *Colour Perception: Mind and the Physical World*, pages 381–430. Oxford University Press, New York.

[Mausfeld and Niederée, 2003] Mausfeld, R. and Niederée, R. (2003). Can a physicalist notion of color provide any insight into the nature of color perception? *Behavioral and Brain Sciences*, 26(1):41–42.

[McGinn, 1996] McGinn, C. (1996). Another look at color. *The Journal of Philosophy*, 93(11):537–553.

[Neumeyer, 1998] Neumeyer, C. (1998). Comparative aspects of color constancy. In Walsh, V. and Kulikowski, J., editors, *Perceptual Constancy: Why Things Look as They Do*, pages 323–351. Cambridge University Press, Cambridge.

[Noë, 2004] Noë, A. (2004). *Action in Perception*. MIT Press, Cambridge, Massachusetts.

[Noë, 2006] Noë, A. (2006). Experience without the head. In Gendler, T. S. and Hawthorne, J., editors, *Perceptual Experience*, pages 411–433. Oxford University Press, New York.

[Palmer, 1999] Palmer, S. E. (1999). *Vision Science: Photons to Phenomenology*. MIT Press, Cambridge, Massachusetts.

[Pylyshyn, 1999] Pylyshyn, Z. W. (1999). Is vision continuous with cognition? the case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3):341–423.

[Riggs and Peterson, 2000] Riggs, K. J. and Peterson, D. M. (2000). Counterfactual thinking in pre-school children: Mental state and causal inferences. In Peterson, D. M. and Riggs, K. J., editors, *Children's Reasoning and the Mind*, pages 87–99. Psychology Press, Hove, UK.

[Riggs et al., 1998] Riggs, K. J., Peterson, D. M., and Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, 13:73–91.

[Rüttiger et al., 1999] Rüttiger, L., Braun, D. I., Gegenfurtner, K. R., Petersen, D., Schönle, P., and Sharpe, L. T. (1999). Selective color constancy deficits after circumscribed unilateral brain lesions. *The Journal of Neuroscience*, 19(8):3094–3106.

[Shope, 1978] Shope, R. K. (1978). The conditional fallacy in contemporary philosophy. *The Journal of Philosophy*, LXXV(8):397–413.

[Stoerig, 1998] Stoerig, P. (1998). Wavelength information processing versus color perception:evidence from blindsight and color-blind sight. In Backhaus, W. G. K., Gliegl, R., and Werner, J. S., editors, *Color Vision: Perspectives From Different Disciplines*, pages 131–147. Walter de Gruyter, Berlin.

[Thompson, 2006] Thompson, B. (2006). Colour constancy and Russellian representationalism. *Australasian Journal of Philosophy*, 84(1):75–94.

[Troost, 1998] Troost, J. M. (1998). Empirical studies in color constancy. In Walsh, V. and Kulikowski, J., editors, *Perceptual Constancy: Why Things Look as They Do*, pages 262–282. Cambridge University Press, Cambridge.

[Troost and deWeert, 1991] Troost, J. M. and deWeert, C. M. M. (1991). Naming versus matching in color constancy. *Perception & Psychophysics*, 50:591–602.

[Tye, 2000] Tye, M. (2000). *Consciousness, Color, and Content*. MIT Press, Cambridge, Massachusetts.

[Valberg and Lange-Malecki, 1990] Valberg, A. and Lange-Malecki, B. (1990). 'colour constancy' in Mondrian patterns: A partial cancellation of physical chromaticity shifts by simultaneous contrast. *Vision Research*, 30(3):371–380.

[Walton, 1984] Walton, K. (1984). Transparent pictures: On the nature of photographic realism. *Critical Inquiry*, 11:246–276.

[Wandell, 1989] Wandell, B. A. (1989). Color constancy and the natural image. *Physica Scripta*, 39:187–192. Reprinted in [Byrne and Hilbert, 1997b], 161–175.

[Wyszecki and Stiles, 1982] Wyszecki, G. and Stiles, W. S. (1982). *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley, New York. Second edition.

[Zaidi, 1999] Zaidi, Q. (1999). Color and brightness induction: from Mach bands to three-dimensional configurations. In Gegenfurtner, K. R. and Sharpe, L. T., editors, *Color Vision: From Genes to Perception*, pages 317–343. Cambridge University Press, Cambridge.

[Zaidi, 2001] Zaidi, Q. (2001). Color constancy in a rough world. *Color Research and Application*, 26:S192–S200.