

# HOW TO THINK ABOUT PERCEPTUAL PHENOMENAL CHARACTER

by

WILLIAM ALLEN SHARP

A thesis submitted to the University of Birmingham for the degree of  
DOCTOR OF PHILOSOPHY

Department of Philosophy  
School of Philosophy, Theology and Religion  
College of Arts and Law  
University of Birmingham  
March 2022

UNIVERSITY OF  
BIRMINGHAM

**University of Birmingham Research Archive**

**e-theses repository**

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

*For my parents*

## Acknowledgements

I need first to thank Scott Sturgeon and Maja Spener for the countless discussions of the included essays. Their advice to *scratch the granite*, along with probably depthless patience, turned the original versions of these papers into something I never thought I was capable of writing. I need also to thank Henry Taylor for many further discussions, anent especially the third essay, and Sergio Farias de Souza Filho for the same.

I want to thank all my peers in the department of Philosophy at University of Birmingham who read earlier drafts of the included essays or quizzed me over early presentations of the content. In chronological order, they are: Tom Davies, Michael Roberts, Casey Elliot, Alex Miller Tate, Federico Bongiorno, Tom Baker, Jack Kelly, Eugenia Lancellotta, Marco di Natale, John Murphy, Kathleen Murphy-Hollies, Harry Goldborn, Aisha Qadoos, and Jess Sutherland.

Thanks also to psychologist extraordinaire Elif Baştan for the bafflement at these essays' former "creative" style, which bafflement hastened me to get on board with my supervisors' advice to scratch the granite.

I want finally to thank my mom, dad, and brother for always encouraging me throughout my years abroad. I dedicate this book to my parents, without whose constant intellectual influence I would be unwise indeed.

# Contents

<i>Introduction</i>	5
1. A new case for eliminativism	18
2. Disjunctivism and supersaturated red	73
3. Russellian representationalism and the stygian hues	114
4. In defense of the response-independence of color	145
<i>Bibliography</i>	160

## Introduction

Three of the four papers of this thesis are on the topic of perceptual experience, which topic makes up a big part of Philosophy of Mind. Philosophy of Mind is an expansive subject, but two of its most central questions are these: What is the nature of phenomenal character—of, that is, the so-called what-it’s-likeness of experience? And second, what is the nature of intentionality—of, that is, the phenomenon of being of or about things? And perceptual experience, it is typically held, exhibits both phenomenal character and intentionality. So, from the perspective of Philosophy of Mind, it is doubly important.

But one of the main contributions of this thesis is that perceptual experience might not be so important for exhibiting phenomenal character. And this because perceptual experience might not have phenomenal character at all. But before I elaborate on this idea, I need to clarify how exactly I am thinking of perceptual experience. What is perceptual experience?

Perceptual experience carves out a proper sub-category of experience in general. Pinning down what exactly makes for something’s counting as experience, in its most general sense, is a notoriously difficult question. But we can go a long way in defining the notion by pointing out paradigms: moods, like feeling glum or upbeat; pain-sensations, like a throbbing pain in one’s toe, and other bodily sensations, like tickles; emotions, like anger or elation; visual experience, like visual experience of a ripe red tomato; auditory experience, like auditory experience of a soprano’s quavering C<sub>b</sub>; among many others. These are all experiences. But we will be concerned just with experiences that share much more in common with the last of the two paradigms mentioned. These experiences, visual experience of a tomato and auditory experience of a soprano singing, are perceptual experiences. And perceptual experience, though also tricky to define, is not as hard to pin down as experience-in-general.

To home in on the conception of perceptual experience I prefer, we need first to specify *sense-modalities*. Human sense-modalities (or as they are more colloquially called, the senses) include sight, touch, hearing, smell, and taste—and possibly many more besides, like the vestibular sense (elsewhere “equilibrioception”, though colloquially “balance”), proprioception (the sense of the position of the parts of one’s body), and others. Any of these sense-modalities counts as perceptual, for me, only if there is a domain of external properties with which the modality puts subjects in contact. And an experience, for me, counts as perceptual only if it is an experience in one of the perceptual modalities.

In this work, I assume that sight (vision), hearing (audition), and touch are three such perceptual modalities. But I do not weigh in on whether taste (gustation) or smell (olfaction), for instance, are also perceptual. I leave that as an open question. If taste and smell do not place experiencing subjects in contact with external properties, like smells of rose or mildew, or tastes of sweetness or bitterness, then they are not perceptual. And if they are not perceptual, then the conclusions I draw in the thesis regarding perceptual experience in general will not bear on how we ought to conceive of experiences that fall into those modalities. As I said above, one of what I take to be the major contributions of this thesis is that perceptual experience does not have phenomenal character. If taste and smell are not perceptual, then I will not have argued that they do not have phenomenal character. For all I will argue, it may well be that tastes and smells just are phenomenal properties (i.e., those properties collections of which are phenomenal character).

Following on that point, because I am concerned just with perceptual experience, the conclusions I reach in this work will not straightforwardly bear on how we should think of experience in general. I will argue that perceptual experience lacks phenomenal character, but I will not comment on whether we should think that experience in general lacks phenomenal character. For all I will show, it may well be that non-perceptual and non-sensory experiences, like feelings of boredom or excitement, are also just phenomenal properties.

So the experiences with which I will be principally concerned are visual, auditory, and tactile. (And yet, following what seems to be a potentially regrettable tradition in the literature on perception, most of the emphasis will be on vision.) And I focus on experiences falling into these modalities because they are perceptual, which means that to each modality there corresponds a domain of external properties for subjects' experiences to be of or about. What are these external properties? They are nothing other than those of the manifest image: things like color, motion, shape, timbre, pitch, volume, pressure, among others.

I have emphasized twice now that it is specifically perceptual experience that I will be investigating. But it also deserves special emphasis that it is perceptual *experience* that I will be investigating. The topic of my concern is not perceptions which are not experiences. And, so, I will not be weighing in on the very interesting recent debates about whether unconscious perceptions mark out a coherent category (see Block, 2016; Phillips, 2016; Block and Phillips, 2016; Taylor, 2020). That said, I will comment later in the introduction on how I think my conclusions bear on that topic.

I said perceptual experience is interesting from the perspective of Philosophy of Mind for exhibiting phenomenal character and for exhibiting intentionality. I should say now what these features of perceptual experience intuitively amount to. The phenomenal character of an experience is what it is like to undergo that experience. And, so, we say that the phenomenal character of visual experience of a ripe, red tomato is what it is like for the subject to undergo visual experience of a ripe, red tomato; and the phenomenal character of auditory experience of a soprano's quavering C<sub>b</sub> is what it is like to undergo that auditory experience. What it is like to undergo these experiences, then, makes up features of the experience, its what-it's-like features, on which subjects are supposed to be able to train their awareness—or, *introspect*.

Why should phenomenal character be thought interesting? Most interest in it, at least since the 1970s, seems to be a product of the perceived difficulty of accommodating phenomenal



character within a metaphysics that is increasingly scientific—i.e., within a worldview intent on explicating all of reality in terms amenable to the physical sciences. I date recent interest in this puzzling aspect of phenomenal character to the 1970s because it was in that decade that Saul Kripke delivered the famous lectures that would later become his *Naming and Necessity* and that Thomas Nagel's seminal article "What is it like to be a bat?" was published. Both Kripke's book and Nagel's article highlighted the stark obstacles to be faced by the increasingly scientific worldview that was dominating Philosophy at the time, and which remains popular today.

Now, I do not want to overstate Kripke and Nagel's roles in fomenting philosophic interest in what perceptual experience is like. Though preoccupation explicitly with what perceptual experience is like is a relatively recent phenomenon, philosophic interest in the apparent incommensurability of physical explanation and experience more generally dates much further back than that. A sense of this incommensurability was presaged centuries earlier by Descartes, who famously argued for the distinctness of so-called *thinking substance* (of which minds are paradigm instances) and *extended substance* (basically, matter). So, phenomenal character is interesting its apparent incommensurability with physical explanation. But if you ask certain philosophers, you might also get the response that it is not only interesting for the above, but interesting for being precisely the stuff which makes life worth living (Dennett, 2005, p. 92 attributes this sentiment to Wilfred Sellars).

Let us move on to perceptual experience's intentional aspects. Intentionality, I have said, is the phenomenon of being of or about something. So, for perceptual experience to be intentional is for it to be of or about something. Intuitively, that "something" is the external world (by which I just mean the world outside the subject). Perceptual experiences are interesting because they afford subjects opportunities to have the external world presented to them, or because they afford subjects opportunities to represent the external world.

One might consider the intentionality of perceptual experience interesting for a few reasons. We might find intentionality simply intrinsically interesting. Perhaps we take it that it is simply a puzzling feature of reality that some of its constituents are of or about some other of its constituents. (If you are not convinced that that, at first blush, is a real problem at all, consider this nearby variant of the question: How is it that anything, to the exclusion of some other things, could be or about anything at all? After all, it is deeply intuitive to think that not everything is of/about something. Why are some things of/about things while other things are not?) Or perhaps we find it more puzzling when certain of those constituents are physical. And, so, if one thinks that human subjects are beings that are physical through and through, then one might wonder how human subjects, but not rocks, stars, rivers, and the like, manage to enter into states (experiences) that are of or about other things. Philosophers engaged in so-called semantic naturalization projects try to make sense of these sorts of questions. Alternatively, we might find intentionality not intrinsically interesting but interesting for how it connects up with other issues in Philosophy. Such issues might include for instance the epistemic contact perceptual experience affords subjects with the world: the perceptual modalities afford subjects knowledge of things, like knowledge of the perceivable character of the external world (cf. Campbell, 2002), or knowledge of external objects' instantiating perceivable qualities (Brewer, 2011). And perhaps there are further areas of Philosophy with which perceptual experience connects that makes perceptual experience worth investigating: like, to give one example, issues in Philosophy of Action, with phenomena like agency and responsibility.

Before moving on to whether we are right to find perceptual experience interesting for exhibiting phenomenal character and intentionality, I want to return to two points made above. I said that we might find it puzzling how physical things, like human subjects, exhibit intentionality. And I said we might be puzzled by how phenomenal character should fit into the increasingly physical-scientific picture we are getting of the world. It needs noting that the current work is firmly in the tradition of *physicalist* explorations of perceptual experience.

Physicalism, very roughly, is the thesis that everything is physical. And something is physical, also very roughly, if it is expressible either directly or derivatively in terms of the posits of fundamental physics (think quantum field theory). And, so, a physicalist theory of perceptual experience will be one which holds that perceptual experience can be accounted for in terms which comport with those of the physical sciences. For an excellent overview of why theorists of perceptual experience should be committed to physicalism, see David Papineau's unmissable "The Rise of Physicalism" (2000).

We should be able to leave our characterization of physicalism roughly put because commitment to physicalism does not play an explicit role in any of the arguments of the thesis. But the commitment needs noting anyway for the following two reasons. It needs noting, first, because it may be that there is a deep-going bias towards physicalism that guides much of the reasoning of this work in a way that I have failed to notice. But the more important reason for noting it is this. Commitment to physicalism plays a role in filtering which perceptual theses I think are most worth investigating. As we will shortly see, two of my papers are dedicated (respectively) to examining the prospects of fundamentally characterizing hallucinatory experience in terms of indiscriminability from veridical perceptions, and to examining the prospects of accounting for perceptual experience in terms of representational contents that are constituted by external objects and properties. Again, I think it is worth spending the time examining these theses' bona fides because they are the most promising physicalist accounts of perceptual experience.

I return now to the reasons for thinking perceptual experience is philosophically interesting. If the contention of my main contribution is right, then perceptual experience is not interesting for exhibiting phenomenal character. This because, again, perceptual experience does not have phenomenal character (I will argue). This does not bear on whether perceptual experience should be thought interesting for exhibiting intentionality. For all I will say, perceptual experience is special precisely because of its intentional aspects. And that is one of the things I think the

papers of this thesis do suggest. One of the things they can be taken to suggest is that our theorizing about perceptual experience should be dedicated to making sense of its intentional aspects. But we will return to this point after I introduce the main paper of the thesis.

Before I summarize the thesis's main contribution, one final aspect of perceptual experience needs to be introduced. Familiarly, there are instances of perceptual experience which (to put it roughly) present or represent the external world as being otherwise than it is. I am meaning here cases of hallucination and illusion. In the literature with which I will be engaged, those experiences which fail in respect of presenting or representing the world as it is we label *Bad*. Hallucinations and (arguably some) illusions count as such *Bad* cases. And those perceptual experiences which succeed in presenting or representing the world as it is, we label *Good*. These labels divide particular experiences extremely coarsely, into those which get things right and those which do not. But almost all experiences are liable to get things a bit wrong. And we do not want to label every experience therefore *Bad*. Moreover, probably very many experiences which do strike us as obviously *Bad*, like a hallucination of a talking dog, can still manage to get some things right, like various aspects of the scene in which the hallucinated object seems to be situated. But we nevertheless will probably not be inclined to call these experiences therefore *Good*. It is probably best, then, to say that Goodness and Badness occupy ends of a spectrum, and that most experiences land on this spectrum away from the extremes (cf. Sturgeon, 2008, pp. 113-115). That said, we can speak of respects in which experiences are *Bad* (or *Good*). Certain experiences we can still say are hallucinatory in respect of involving the presentation or representation of an object which is not before the subject. And, so, if we speak of a visual *Bad* case of a dog, we will take this to mean just that a dog is hallucinated, but not that everything about the presented or represented scene is inappropriate given what is really before the subject. We only need to understand that a *Bad* case of such-and-such is not necessarily altogether *Bad*. With this understanding of the extents to which perceptual experiences may be *Good* or *Bad*, we can turn now to the summary of the main contribution of the thesis.

The longest of the papers—the flagship paper—argues for *eliminativism* with respect to perceptual phenomenal character. This is the view that perceptual experience has no (does not instantiate) phenomenal character. The argument works in stages, where each progressive stage challenges the legitimacy of different sets of evidence we might have thought we had for believing that there is phenomenal character in the Good case. I conclude the first stage of the argument by saying that we have no reason to believe there is phenomenal character in the Good case. From there, in the second stage, I explain why I think this conclusion leaves us in a dialectic position which allows for a much easier time also rejecting there is phenomenal character in the Bad case. I conclude this second stage with saying we should reject that the Good case has phenomenal character and also reject that the Bad case has phenomenal character. This constitutes an out-and-out eliminativism with respect to perceptual phenomenal character.

What is the value of eliminating perceptual phenomenal character? Well, if perceptual experience does not instantiate phenomenal character, then much of philosophers' theoretic efforts of the last several decades have been misplaced. A practical consequence of eliminativism is that our theoretic interests are better directed elsewhere. I said above that if perceptual experience does not instantiate phenomenal character, then we should concentrate on the other aspect of perceptual experience which makes it philosophically interesting, namely, its intentionality. Intentionality, we saw above, is interesting for a host of reasons. Accordingly, by eliminating perceptual experience's phenomenal character, we are prompted to place our theoretic efforts in illuminating how perceptual experience's intentionality affords these epistemic connections, and how it can be accommodated in terms amenable to the physical sciences. (For an excellent treatment of this latter issue, see Nicholas Shea's, 2018, *Representation in Cognitive Science*.)

And I want to stress that by cutting off a major source of perceptual experience's philosophic appeal, its alleged phenomenal character, we do not thereby diminish the phenomenon. Perceptual experience's intentionality is, at first blush, distinct from, say, the intentionality

exhibited by thought. Thought, or thinking, is also of or about things. Moreover, like perceptual experience, thought is also of or about the external world. But there is an apparent difference between the intentionality exhibited by thinking, on the one hand, and perceptual experience, on the other. If the thesis's flagship paper is wrong, we might chock this difference up to perceptual experience's having phenomenal character and thought's not having phenomenal character. But if we eliminate perceptual phenomenal character, as I argue we should, then we will have to explain the difference in another way. Thus, we have a putatively more difficult puzzle on our hands in the case that we eliminate perceptual phenomenal character: how to make sense of the difference between the intentionality exhibited by perceptual experience and the intentionality exhibited by thought if not by perceptual phenomenal character.

I have said that eliminativism is practically important because it steers our theoretic interests to better havens. But this is not only true for philosophers. Those *scientists* inclined to use philosophic jargon, or who think of themselves as investigating which mechanisms underpin perceptual phenomenal properties: they too could benefit from having their quarries replaced with something else instead.

Crucially, the practical upshot of eliminativism is not that it is a waste of time to theorize about perceptual experience. No, perceptual experience still exhibits philosophically interesting features—and many. For instance, it will still be appropriate to ask how we ought fundamentally to characterize perceptual experience? And this question brings us to the other two papers of the thesis that are also on the topic of perceptual phenomenal character.

The first of these papers argues that the manner of characterizing hallucinatory experience that is popular especially in the UK, as consisting in subjects' being unable to know hallucinations distinct from corresponding Good cases, has not yet been shown to be able to accommodate a type of afterimage-experience. The troublesome afterimage-experience is experience as of a red patch more saturated than a fully saturated red surround. Saturations beyond maximum are

impossible. So the experience intuitively involves the presentation of an impossible saturation-property. This is a problem for the characterization of the Bad case on which the phenomenal character of hallucinatory experience consists just in inability to tell experiences apart from corresponding Good cases, for there are no corresponding Good cases here.

This paper is of independent interest for contributing to the debate on how we should characterize the Bad case—and perceptual experience in general. Hallucinatory experience may be interesting in its own right. But the conclusion of the paper is relevant for any philosopher interested in examining the nature of perceptual experience generally. If one important sub-category of perceptual experience, the Bad case, does not consist in subjects' inability to know the experiences to be distinct from certain Good cases, then neither does perceptual experience in general consist in this. Investigating hallucination, we see, sheds light on how we should think about perceptual experience generally. The fundamental characterization of perceptual experience is not to be given in terms of indiscriminability from Good cases. We must look elsewhere for that characterization. And, so, the second paper serves the role of giving advice regarding how we characterize perceptual experience, irrespective of whether eliminativism is right.

The final of the papers on the topic of perceptual phenomenal character, like the second paper, advises how we should think about perceptual experience even in the case that perceptual phenomenal character does not exist. In the third paper, I argue that the most popular externalist representational theory of perceptual experience, Russellian representationalism, faces counterexamples in the so-called *stygian* afterimage-experiences. This version of representationalism sees experiences as supervening on contents, and contents, on the view, are constituted by the external objects and properties the experiences represent. (Because experiences supervene on items constituted by external objects/properties, the view is externalist. It is representationalist for positing experience's supervenience on representational

contents.) The stygian afterimage-experiences present a problem for Russellian representationalism, because the view assigns each distinct experience identical contents, namely, contents missing identical bits of constituting structure. This happens because each stygian afterimage-experience is an experience as of an impossible color property. Each of the experiences represents surfaces as being dark-as-black and yet hued. But such experiences are not of possible color properties. To get a very rough sense of why, consider that maximally dark surfaces are just black ones, and that hued surfaces are not black. So, there are no possible surfaces possible that are simultaneously dark-as-black and hued. I argue in that paper that, due to other commitments proponents of Russellian representationalism are wont to make, this makes the properties unavailable as constituents of the relevant contents. And the experiences' associated contents consequently end up being identical for missing identical bits of structure. (Each experience would have as its content, "*This surface is...*," where the ellipsis denotes missing structure.) The conclusion of this paper is important because Russellian representationalism is probably the most popular perceptual thesis in the world today.

Before I introduce the final paper, I want briefly to return to the matter of the papers' lack of engagement with the literature on unconscious perceptions. The conclusion of the main contribution of the thesis has consequences for that debate. I argue in that paper that there is no perceptual phenomenal character. Now suppose, along with what I take it is the majority, that *unconscious perception* marks out a genuine category of mental event. There are unconscious perceptions. In the case that perceptual phenomenal character does not exist but unconscious perceptions do, then the difference between conscious and unconscious perceptions cannot consist in conscious perceptions, but not unconscious perceptions, exhibiting phenomenal character. This cannot, I mean, be what is meant by the distinction. The difference between unconscious and conscious perceptions will have to be cashed out in terms of something else. What that something else is, however, so long as it is not phenomenal character, I leave to involved parties.



I turn now to the final, very short, paper of the thesis. The final paper may at first blush appear tangential to the others. Each of the others is about how we should think about perceptual phenomenal character. But the final paper is on the topic of the nature of color. Let me briefly say how that paper fits with the rest.

The main paper of this thesis argues for the elimination of perceptual phenomenal character. I have said that an important question to ask after eliminativism is embraced is What fundamental characterization should we give to perceptual experience? In answering that question, this further question feels, for me, unavoidable: What are the qualities we are aware of in perception? (For more on the relation of this question to the more familiar questions of perceptual phenomenal character's grounds, see the Afterword following the first paper.) Because this question is raised by the eliminativism on offer, and because the bulk of this thesis takes up with the question specifically of how we should talk about the phenomenal character distinctive of experience of color, a pressing specification of the above question is: What is color?

In the final paper, I respond to a pair of recent arguments against the family of views committed to colors' natures being individuable independently of features of subjects. On the view I defend, redness, greenness, and all the rest of the colors are external properties of objects which would be just as perception reveals them to be even were all subjects suddenly to cease to be, or even had no subject existed at all.

I hope that this last paper is a breath of fresh air given the negative character of the rest of the thesis. The last is the only paper in which I *defend* a view. The rest of the papers see me arguing for the falsity of views, or for the misguidedness of entire research programs.

#### *Order of the papers*

As I have said, the papers on the topic of how we should characterize Bad experience are written in a language intended to make them interesting to the broadest audience. To achieve that, those

two papers suppose realism about phenomenal character. This makes it so that there is no best way to order the papers. If I had started with the paper on eliminativism, the two papers on the Bad case would read jarringly. Or if I had started with the papers on (the phenomenal character of) the Bad case, then on reading the eliminativism paper, my reader may be inclined to wonder why in the world I spent two papers discussing how to characterize the phenomenal character of the Bad case if we should not think of it as existing.

In the end, I decided to put the eliminativism paper before the other two. I did this because I thought it would be better for my reader to have a slightly jarring experience rather than that she be forced to wonder why on earth I made the choices I did.

## A new case for eliminativism

**Abstract:** In this paper I make a case for eliminating perceptual phenomenal character. The case proceeds in two stages. The first is an argument for the conclusion that we should not believe there is phenomenal character in the so-called Good case of perceptual experience. The argument is substantively premised on two ideas. First, if we appeal only to features of the Good case, there is no uncontentious way to steer the uninitiated to the phenomenal character of their experience. Second, the reasons for this carry over to the Bad case too, so concerning the Good case there is no uncontentious way to steer the uninitiated to the phenomenal character of their experience.

The second stage is a reflection on where the first stage leaves us. There I argue that considerations of motivation and parsimony encourage rejecting that there is phenomenal character in the Bad case too.

“The riddle was: why couldn’t we live in the mind.

The answer was: the barrier of the earth intervened.”

-from “Prism” by Louise Glück

Eliminativism, the idea that phenomenal character is never actually instantiated, is gaining in popularity lately. Strong illusionism, the view that phenomenal character is an illusion and is never actually instantiated, is today being championed by a number of philosophers, among them Dan Dennett (2016), Keith Frankish (2016), Derk Pereboom (2016), and Francois Kammerer (2018). However, strong illusionists have a penchant for construing ‘phenomenal character’ parochially—as denoting qualia, the intrinsic, non-relational, ineffable, subjective qualities of experience. Doing this limits the appeal of strong illusionism.

The goal of the present paper is to investigate the prospects of eliminating phenomenal character as minimally construed. In one respect, then, this paper is ambitious. Strong illusionism is an eliminativism about something few philosophers today believe in. By defending in this paper the elimination of phenomenal character minimally construed, we will be saying that something most philosophers believe in does not exist.

But in another respect, the goal of this paper is modest: it is to develop a strategy for making a case just for *perceptual* phenomenal character’s elimination. What will be argued is that we should reject that perceptual experience ever instantiates phenomenal properties. No proposal will be made in defense of eliminating phenomenal character in the case of, say, pain, or doldrums. And when is an experience perceptual? I will be assuming that an experience is perceptual exactly when it belongs to a sensory modality for which there exist external qualities to be of, or about. I

will be taking for granted that vision, audition, and touch are perceptual in this sense. These assumptions I make just to carve out a subject matter, but the argument does not depend on it.

Finally, this paper is not a defense of any sort of illusionism. The strategy pursued below, if anything, makes of phenomenal character more a myth than an illusion. None of us is undergoing an illusion that phenomenal character exists. Rather, the idea of perceptual phenomenal character, at least as instantiated by veridical perceptual experience, is one we have inherited despite never having been in possession of a good reason to believe it exists.

The case to be made for eliminativism proceeds in two stages, with the first making up the bulk of the paper. The first stage involves arguing for the conclusion that we should reject that there is phenomenal character in those cases in which we perceive the world for how it is—in the so-called Good case. That argument (section 2) is substantively premised on two ideas, to be defended in turn. The first (section 3) is that realists about phenomenal character will not be able to give newcomers to Philosophy of Mind a theory-independent reason to believe that the Good case has phenomenal character, if that reason avoids cases that are not Good—i.e., so-called Bad cases, cases of perceptual experience where it is not the case that one perceives the world for how it is. The second (section 4) is that, if the first premise is correct, then neither will the realist be able to give newcomers a theory-independent reason to believe there is phenomenal character in the Good case which reason is drawn itself from the Bad case. These substantive premises spell bad news for realism if we assume phenomenal character must be specifiable independently of theory in order that the enterprise of theorizing about it be justified (section 2.1).

The argument depends on three further premises (section 5). First, our reason to believe the Good case has phenomenal character must be drawn from the Good or Bad case; second, if no novice can have a reason to believe the Good case has phenomenal character, then no one can;

and third, parsimony prescribes rejecting the existence of property-types no one has a reason to believe in. The five premises together entail that we should reject that there is phenomenal character in the Good case.

This is a strange place to end up. It apparently leaves us with a view on which there is phenomenal character only in the Bad case, which is a view no one has ever argued for and one that would in any case attract very few realists. This brings us to the second stage (section 6). The second stage canvasses a choice point. Do we embrace a view that has never been argued for and that posits more properties, or do we eliminate phenomenal character in the Bad case too? I argue that considerations of motivation and parsimony encourage elimination. This gives us the elimination of perceptual phenomenal character, minimally construed, altogether.

### **1. What exactly is being eliminated?**

'Phenomenal character' is the mass noun (phrase) which designates the phenomenal properties instantiated by an experience. So let us start with phenomenal properties. Classically, phenomenal properties were the properties of sense-data (cf. Place, 1956, p. 49), sense-data being the mind-dependent objects of subjects' immediate awareness (Huemer, 2011).<sup>1</sup>

---

<sup>1</sup> I follow Huemer's SEP characterization here. Huemer does not fail to note, however, that sense-data, as the expression was originally used by *fin de siècle* Analytic philosophers, were more neutrally characterized, as denoting neither mind-dependence nor mind-independence. But because the cousin of 'phenomenal properties', 'qualia', in its modern sense coined in the late 1920s, was originally intended to denote the properties of mind-*dependent* sense-data (Frankish, 2012, p. 667), and because U.T. Place, three decades later, writes of phenomenal properties as being the properties of the "mythological 'objects' in the mythological 'phenomenal field'" (Place, 1956, p. 49)—which sounds like a synonymous manner of picking out sense-data—I will gloss over the neutral *fin de siècle* usage. Less long-windedly: because by the time the expression 'phenomenal properties' was in currency phenomenal properties

The phenomenal properties of sense-data were traditionally described as ineffable, non-complex, intrinsic, private, and subjective (Dennett, 1988; Frankish, 2012). And, so, our very-first-pass characterization of phenomenal properties will describe them as being the ineffable, intrinsic, simple, private, subjective properties of sense-data. Phenomenal properties thus construed will not be the target of our elimination, because those properties, for most intents and purposes, have already been eliminated. Only a small cohort of philosophers of the past thirty years (Foster, 1990; Robinson, 1994; O’Shaughnessy, 2003) have posited sense-data.

Dropping the requirement that phenomenal properties be instantiated by sense-data, we are left with a conception of phenomenal properties whereunder they are ineffable, intrinsic, simple, private, subjective properties *of experience* (or of subjects). Such a characterization we might still call classical (cf. Frankish, 2012, throughout).

Phenomenal properties so construed will be part of the target of the present elimination. Not many philosophers—not those, anyway, who are inclined to accept physicalism (Frankish, 2012, p. 668)—today take it that the explicanda of their theorizing are phenomenal properties thus construed. For many intents and purposes, those properties too have already been eliminated. Only a smallish cohort of contemporary philosophers posit these properties.<sup>2</sup> Today, it is more common to drop the requirements of intrinsicness, ineffability, simplicity, privacy. Phenomenal properties may be those things, but it is not built into the notion that they be.

---

were thought to be properties of mind-dependent sense-data, I will treat the expression as having historically been applied to mind-dependent sense-data.

<sup>2</sup> Nevertheless, and perhaps disappointingly, phenomenal properties so construed constitute the target of the eliminativism of Dennett (1988), (2016), Frankish (2016), Pereboom (2016), and Kammerer (2018)—the Strong Illusionists.

In the face of the classic requirements' being dropped, the obvious question is, 'Then what *are* phenomenal properties?' because now there is basically nothing left to positively characterize them. Since Thomas Nagel's seminal (1974) essay, it has been popular to adopt the following locution in our characterizations of 'phenomenal character': phenomenal character is what-it's-likeness. The phenomenal character of a visual experience as of red is what it is like to have a visual experience as of red; the phenomenal character of an aural experience as of C# is what it is like to aurally experience C#. On this construal, which drops all of the classic requirements save for subjectivity, phenomenal properties are those properties individuated by what it is like to have them (cf. Chalmers, 2018, p. 6).<sup>3</sup>

I borrow this notion from Chalmers (2018). But Chalmers suggests (in correspondence) that this characterization is meant to capture that phenomenal properties are fully individuated with reference to what it's like to experience something. And their being fully individuated with reference to what it's like is meant, moreover, to accommodate that phenomenal properties just are what-it's-likeness. And, so, on this characterization we can say that *phenomenal red* just is what it's like to see red; *phenomenal Gb* just is what it's like to hear Gb; and so on.

Before answering whether phenomenal properties so construed will be the target of the present elimination, it is worth looking at a fuller, alternate contemporary characterization of phenomenal character construed as what-it's-likeness. Here is a particularly illustrative homing-in on the (alleged) phenomenon from Block (1995):

---

<sup>3</sup> Note, however, that this conception leaves open that what constitutes what an experience is like are classic phenomenal properties.



“[H]ow should we point to P[henomenal]-consciousness? One way is with rough synonyms. As I said, P-consciousness is experience. P-consciousness properties are experiential ones. P-conscious states are experiential, that is, a state is P-conscious if it has experiential properties. The totality of the experiential properties of a state are "what it is like" to have it. Moving from synonyms to examples, we have P-conscious states when we see, hear, smell, taste, and have pains. P-conscious properties include the experiential properties of sensations, feelings, and perceptions, but I would also include thoughts, desires, and emotions. A feature of P-consciousness that is often missed is that differences in intentional content often make a P-conscious difference. What it is like to hear a sound as coming from the left differs from what it is like to hear a sound as coming from the right. P-consciousness is often representational”  
(230).<sup>4</sup>

I offer this alternate characterization because Block (1995) is seminal, and because it is illustrative to note the differences between it and the more straightforward-seeming characterization given just before it. First, note that Block construes what-it’s-like as a matter of sum totals of phenomenal properties. Phenomenal properties make up, then, what-it’s-likeness rather than being individuated by what having them is like. It is compatible with phenomenal properties’ making up what it is like to undergo a given experience that they be individuated otherwise than by what having them is like. An analogy should make this clear. Suppose a populace is identical to an aggregate of individual citizens (in the way that what-it’s-likeness is identical to phenomenal character, which here is just the mass noun–phrase which designates aggregates of phenomenal properties). This is compatible with individual citizens’ being individuated by the populace to which they belong (or that they partly comprise). But it is far from guaranteed that individual

---

<sup>4</sup> If the reader feels that this characterization begs the question against views which take differences in phenomenal character to *make for* differences in intentionality, rather than the reverse, she can read Block’s “makes for” of the quotation’s penultimate sentence as “accompanies” and nothing should be lost.

citizens be individuated like that. If anything, it is much more likely that the individual citizens belonging to, or comprising, the populace be individuated *not* by reference to the populace they belong to or that they comprise.

We can extract from this analogy that the synonymy of ‘phenomenal character’ and ‘what-it’s-likeness’ is no bar to phenomenal properties’ being individuated otherwise than by reference to what it is like to have them. This is to say that we can pull Block’s characterization of phenomenal character apart from (what I will call) Chalmers’s. On Chalmers’s characterization, phenomenal properties are just those properties individuated by what it is like to have them (see also Chalmers, 2018, fn. 2 for this usage). On Block’s characterization, phenomenal properties might be individuated otherwise than by reference to what it is like to have them, even though they nevertheless aggregate into what-it’s-likeness.

And, so, if the quotation of Block’s is not strictly a fleshing out of the Chalmersian construal of ‘phenomenal character’, then we should consider it a separate construal. With that said, however, I am going to take interpretative license and make the simplifying assumption that Block’s “experiential properties” just are Chalmers’s phenomenal properties. Phenomenal properties make up phenomenal character (given the mass-noun functioning of ‘phenomenal character’) the way experiential properties make up what it’s likeness on Block’s elaboration. So far as I have been able to work out, nothing is lost in doing this.

Second, although the above quotation of Block’s is intended to get us to home in on phenomenal character by way of providing synonyms, there is a danger that some of those synonyms will lead the uninitiated to properties distinct from phenomenal ones. Phenomenal properties, Block says, are experiential properties, and the sum total of an experience’s experiential properties make up what the experience is like. But it is not all of an experience’s properties that contribute to what-it’s-likeness. Plausibly, an experience’s having happened at 4

o'clock, or its having taken place at such and such longitude, are properties of the experience that do not contribute to experience's what-it's-likeness. It is plausible, in other words, that those two properties are not phenomenal properties. What is more plausible is that it is only a proper subset of an experience's properties which contribute to what it is like to have it. In other words, plausibly it is only a proper subset of an experience's properties which are its phenomenal properties. With this caveat in place, the defense of our argument's main premise will run much more smoothly.

And finally, one assumption I will be making needs to be brought to the fore. So far in our brief look at the classic and contemporary construals of 'phenomenal properties', we have seen that the *subjectivity* of phenomenal properties was never abandoned. Accordingly, I will treat subjectivity as being built into the notion of phenomenal character. Phenomenal properties are, at least partly, subjective properties. By 'subjective', I just mean subject-involving. (And 'subject' I think we should be able to leave unanalyzed.) This means that things like spectral surface reflectances, for instance, which are not subject-involving—object surfaces instantiate these properties independently of any subjects' existing—cannot count as phenomenal properties.

Phenomenal properties construed *à la* Chalmers (or Block, if we grant my interpretative assumption) will be the target of our elimination. In the next section, I will begin to make the case for eliminating phenomenal character minimally construed.

## **2. The argument**

Before we can get to the argument, a handful of terms need to be introduced. First, for our purposes a perceptual experience is one of the *Good cases* iff it is an occasion on which the subject perceives the world for how it is. And, so, to use an example made popular by Heather Logue (2013), when a subject visually perceives a yellow, crescent-shaped banana for the yellow,

crescent-shaped banana that it is, we say that that subject's perception of the banana is Good. And for our purposes a perceptual experience is one of the *Bad cases* iff it is an occasion of the subject's failing to perceive the world for how it is, as in hallucination and illusions. And because it will help cut down on verbiage later, let us call phenomenal character instantiated by the Good case *Good character* and phenomenal character instantiated by the Bad case *Bad character*.

Next, let us call a reason a *Bad reason* iff it is explicitly drawn from consideration of Bad cases. And, so, a subject has a Bad reason to believe hallucinations are indistinguishable from perceptions if the belief is explicitly based on reflection on hallucinations she has undergone. And a reason is a *Good reason* iff it is explicitly drawn from consideration exclusively of Good cases.<sup>5</sup> And, so, a subject has a Good reason to believe perceptions are of or about the world if the belief is explicitly drawn from reflection exclusively on Good cases.

And finally, by *novice* I mean a competent thinker not yet familiarized with Philosophy of Mind. Below we will take our novice to be a philosopher.

Here is the argument:

P1. No novice can be given a Good reason to believe Good character exists. (premise)

P2. If no novice can be given a Good reason to believe Good character exists, then no novice can be given a Bad reason to believe Good character exists, either. (premise)

P3. If no novice can be given a Good or Bad reason to believe Good character exists, then no novice can be given a reason simpliciter to believe Good character exists. (premise)

C1. Hence, no novice can be given a reason simpliciter to believe Good character exists. (from 1, 2, 3)

---

<sup>5</sup> It is important to characterize Good and Bad reasons in terms of explicitness because if we did not and it turns out that, say, sufficiently many of reasoners' *priors* are partly shaped by past encounters with Bad cases, then the task of classifying reasons as Good/Bad might be impracticable.

P4. Because we were all novices once, if no novice can be given a reason simpliciter to believe Good character exists, then none of us will ever have been able to be given a reason simpliciter to believe Good character exists. (premise)

C2. Hence, none of us will ever have been able to be given a reason simpliciter to believe Good character exists. (from C1 and 4)

P5. Due to considerations of parsimony, if none of us will ever have been able to be given a reason simpliciter to believe Good character exists, then none of us should believe Good character exists. (premise)

C3. Hence, none of us should believe Good character exists. (from C2 and 5)

The first premise says that the uninitiated cannot be given a reason to believe that there is phenomenal character in the Good case if the reason is explicitly drawn from consideration exclusively of Good cases. This is a substantive premise, and it will take time defending. So in sections 3 - 3.1.1 I will do just that. This first premise, however, would be of limited significance were it not for the second: if no novice can be given a Good reason to believe Good character exists, then neither can he be given a reason to believe it exists which reason is explicitly drawn from consideration of Bad cases. This premise carries considerable weight as well, so section 4 is dedicated to it.

The third premise says that if we can neither find a reason to believe Good character exists by considering the Good case, *nor* find a reason to believe it exists by considering the Bad case, then we will not have any reason at all. This premise and the remaining two will be defended in section 5 below. The fourth premise picks up on our all having been novices once. If a novice cannot be given a reason simpliciter to believe Good character exists, then none of us will have been able to be given a reason. And the final premise is a parsimony assumption. If none of us has a reason simpliciter to believe Good character exists, then we should drop Good character from our ontologies.

### *2.1 Two constraints of theory-neutrality*

Before I defend the argument's substantive premises, I need to introduce two commitments that do heavy lifting in motivating them. The first commitment is a constraint on theory-neutrality that a novice's introduction to Good character must respect. The constraint says there needs to be a theory-independent phenomenon that the various phenomenal theses purport to be in the business of explaining.<sup>6</sup> Our constraint on theory-neutrality says

**Independence Constraint**    What needs to be established is that there is in the first place a phenomenon for theory to be about.

Let me illustrate. On some realist views, Good character is partly grounded in external properties (naïve realism); and on others, Good (and Bad) character is grounded in perceptual experience's representing external properties (representationalism). If a realist can only make it seem to the novice that he can train his awareness on external properties, or on his experience's representing external properties, then she will not have steered him to the appropriate theory-independent stuff. The realist needs to guide the novice to what those resources are called on to account for. *If there is no such thing, then phenomenal theses are in the business of explicating distinct phenomena.* In which case it would be inappropriate to label the distinct phenomena with a single term, namely, 'Good character' (and its cognates).

Another way to motivate the constraint is by appeal to charity. If realists strictly cannot single out Good character independently of calling on theoretic posits, then the posits are apparently being called on to account for themselves. But this stultifies those theoretic enterprises. Charity prescribes taking phenomenal theorists *not* to be doing *that*.

---

<sup>6</sup> Shoemaker (1994) says phenomenal properties are theoretic posits. But his phenomenal properties are not the minimal explicandum introduced in section 1. That said, we will return to his argument in section 4.

## 2.2 Transparency

We will say that perceptual experience is transparent exactly if all of the property-instances (hereinafter “properties” for short) available for attention in a perceptual experience are external, perceptible ones.<sup>7</sup> (By a property’s being external, I mean that it is outside the subject.) This is to say, when subjects introspect their experiences, it will be the only properties available for attention will be external ones.

For illustration, take as an example a visual experience of a ripe Braeburn apple placed on a mahogany wooden table in a uniformly tawny room on an overcast day. This visual experience to be transparent exactly if all of the properties available for attention in introspecting the experience are external ones. The just-mentioned colors, if each available for attention, will be external. Moreover, the sheen and visible texture of the ripe Braeburn, if available for attention, will be external.

By committing to transparency, we are saying that in introspecting the perceptual experience of the scene, if a property is available for the subject’s attention, then it will be external.

Contraposing, we are saying that in introspecting the perception of the scene, if a property is not external, then it will not be available for attention. That is, if a property belongs to the subject’s experience, or to the subject, rather than to the external scene, then it will invariably be unavailable to attention in introspecting the experience.<sup>8</sup> This last claim may look like a hefty one. But *with respect to the Good case* at least, it is not much contested. Perceptual experience’s

---

<sup>7</sup> Cf. Gow (2016): “Perceptual experience is metaphysically transparent if and only if all the properties we are aware of are...externally located” (723).

<sup>8</sup> Strictly speaking, properties will be unavailable for attention even if they are nowhere at all.

transparency is usually objected to by way of counterexample. And those counterexamples usually just involve what are arguably Bad case phenomena, phenomena like blur (Tye, 2003),<sup>9</sup> phosphenes (Block, 1996 pp 34-5), or experienced differences in contrast where no differences in contrast exist (Block, 2003 pp 33-7). Because we will be concerned just with whether we have any Good reasons to believe phenomenal properties exist, we will accordingly prescind from considering these putative counterexamples. (But we will come back to blur in the Objections.)

Now, what are the considerations which positively favor calling perceptual experience transparent? One consideration is transparency's introspective obviousness. Take the above English scene again, the Braeburn on the table in the overcast moor. It is intuitively obvious, we can say, that the only properties available for attention here are not perceptual-experiential ones. Or we might say, as Gil Harman has, "I predict you will find that the only features there to turn your attention to will be features of the presented [scene]" (Harman, 1990 p 667). Of, that is, the colors of the scene, the visible textures of the objects, the sheen of the apple, the mattness of the wood, among other features of the presented scene. This introspective obviousness in the face of a lack of counterexample provides our initial motivation to accept that experience in the Good case is transparent.

In addition to facing no counterexamples and being introspectively obvious, there is a growing body of cognitive neuroscientific evidence which supports it as well. Sauret and Lycan, citing Prinz (2012), Mole (2010), and Sauret (2014), argue that "the leading contemporary cognitive and neurological theories of attention are unanimous in suggesting" that attention has no proprietary inputs, let alone inputs that are mental states (2014 p 365). Instead of having proprietary inputs,

---

<sup>9</sup> It is not obvious that visual blur is strictly speaking Bad. But no explication of what Goodness is would ever advert to blur. We will return to blur in the Objections.



the directedness of attention works the way “bottlenecks, filters and thresholds work; they select objects in virtue of specified properties without representing those properties or the objects that have them” (*ibid.* p 366). That is, when a subject attends to something, her perceptual mechanisms are simply steered or homed, or their outputs are stored in memory (365-66). And perceptual mechanisms represent distal features of subjects’ environments, not mental features. Attention consists just in the homing and steering of those mechanisms and/or the storing of those mechanism’s outputs.

Five years on, Weksler *et al.* (2019) confirm Sauret and Lycan’s point, this time applying the idea to even more mechanisms of attention. Every attention-mechanism recognized today—shrinking receptive fields, raising signal-to-noise ratio, shielding from decay, suppressing distractors, allocating resources—is in the business just of modulating states which represent distal features of the environment (4716-22). (Note that these mechanisms model different aspects of attention and, so, are not in competition, *ibid.* p 4714.) The lesson again is that attention is directed not at anything non-external but involves just the modulation of perceptual processes.<sup>10</sup>

What we have to motivate transparency, then, are its introspective obviousness, a lack of counterexamples, and a growing body of cognitive neuroscientific evidence. Accordingly, we should accept that perceptual experience is transparent.

Before we rely on transparency to make the case for premise 1 of the argument, a caveat is in order. So far, I have only formulated transparency in terms of attention. I have not made the much stronger claim that in perceptual experience the only properties available for *awareness* are external ones (as do, for instance, Tye, 2014 p 40 and Gow, 2016 p 723). That much stronger

---

<sup>10</sup> The findings of Sauret and Lycan and of Weksler *et al.* of course leave open that an attention-mechanism will in future be discovered which mechanism is naturally interpreted as being directed at something experiential. But I will assume my opponent is not interested in anything so hostage to empirical fortune.

claim is arguably false: one of the properties available for awareness in regular cases of introspecting perceptual experience is the experience's property of *being a perceptual experience*. Or, in the specific case of visual experience, we can be aware of the perception's property of *being a visual experience*. Moreover, we can become aware of the experience's property of *being of/about what it is of/about*. So, in perceiving our scene above, we can become aware of the perceptual experience's property of *being an experience of the table-perched Braeburn*.<sup>11</sup> These are properties we can become aware of, and yet they arguably are not external properties. We should acknowledge that there may be very much in the Good case we can become aware of besides external properties. That said, the case to made for our argument's first premise will proceed first in terms of attention. After being framed in terms of attention, we will round out the case by incorporating the observation that attention's limitations notwithstanding there is more we can become aware of in ordinary perceptual experience than just external properties. That there may be more in perceptual experience about which we could become aware besides external properties does not help the realist in her endeavor to provide the novice with a Good reason to believe Good character exists.

### **3. Premise 1**

The first premise of the argument for eliminativism says no novice will ever have been able to be given a Good reason to believe Good character exists. I will begin the case for premise 1 first by way of illustration.

---

<sup>11</sup> Tye (2014) says that these are things about which we can become aware only in a non-de re sense of awareness, which is irrelevant, he says, to establishing the transparency he has in mind (42). Even if Gow (*op. cit.*) and Tye are in the end right to frame transparency as they do, we do not need their much stronger thesis to get the conclusion that none of us will ever have been able to be given a Good reason to believe that perceptual experience instantiates phenomenal character.

Take a novice to the discipline, Novo, who we will suppose is a competent philosopher who has simply yet to get his toes wet in the literature on mind. And take realist about perceptual phenomenal character, Rea. Imagine that Novo, after briefly hearing about perceptual phenomenal properties from a friend, asks Rea to aid in directing his attention to the supposed phenomenal properties of his experience. Perceptual experience's transparency ensures Rea will not be able to get Novo to attend to them. The only properties available for Novo's attention are the external ones of his distal environment. And these are not phenomenal properties.<sup>12</sup>

What we should say, then, is that before having a grip on what phenomenal properties are, no novice will be able to be instructed by a realist to attend to them. This is not to say enough, however, to make the case that no Good reason to believe phenomenal properties exist can be provided to the novice. To do that, more is called for. For starters, we need to incorporate the observation made in the previous section, that although there are no non-external properties available for attention in the Good case, it may be that there are non-external properties available for awareness. The question, then, is whether phenomenal properties are among these. *If* it is the case that no novice will ever have been able to have her *awareness* drawn to phenomenal properties, then we should be able to conclude that no novice will ever have been given a Good reason to believe they exist.

For the remainder of this section, I will present a handful of considerations to motivate accepting that no realist can aid the novice's becoming aware of phenomenal properties, either.

---

<sup>12</sup> *Pave* Tye (2009): "[W]hat it is like to experience red is the same as the color red" (128). Recall that we are assuming phenomenal properties are not exclusively external (section 1). Tye's idea, taken at face value, has the consequence that in a world where no subjects exist, what it is like to experience red is instantiated by Mars.

This will be the *prima facie* grounds for holding no novice could have been given a Good reason to believe perceptual phenomenal character exists.

The question, again, is: Can Rea aid Novo in becoming aware of phenomenal properties, even though there is nothing she can do to direct his attention to them? We imagine now that Rea tells Novo that though Novo cannot attend to the phenomenal properties of his experience, he can nevertheless become aware of them—*by attending* to the properties/objects of his distal environment.

This will not work without considerable further direction on Rea's part because there is very much besides phenomenal properties Novo can become aware of in attending to the external objects/properties of his environment. Chief among the non-phenomenal properties Novo might be aware of in attending to external properties are: the very properties he is attending to. After all, it is usually the case that one is aware of what one attends to. In addition, there are the properties lately noted which kept us from characterizing transparency in terms of awareness: the properties of Novo's experience *being a perceptual experience*, or *being an experience of such-and-such*. But none of the above properties—the properties *being a perceptual experience* and *being an experience of such-and-such*, and the properties attended to—are phenomenal properties. *N.b.* this is not to say that phenomenal properties do not reduce to, or that they are not grounded in, any of those properties. *Before theory*, phenomenal properties are not those. At this point in the dialectic, we cannot appeal to theoretic resources to handle my challenge. What needs to be established is that there even is a phenomenon in need of theoretic accommodation in the first place. This was our Independence Constraint. So, Rea needs to say more than just that Novo can become aware of phenomenal properties by attending to the external properties of his distal environment.

But the difficulty of Rea's task is compounded when we note that were Novo to attend, say, to the red Braeburn (on the multi-colored surround), he would become aware of the colors of the surround. If we placed Novo on a mountain peak and gave him an apple to attend to, he could become aware, in attending to the apple, that his experience is taking place at a high altitude. Novo can also become aware, in attending to the apple, that his experience of the apple is making him hungry or making him nostalgic for his childhood when he snacked often on apples. To give an aural example, were Novo to attend to the timbre of a singing soprano's voice, he would also become aware of the soprano's (or her voice's) volume. But none of these properties are perceptual phenomenal properties. How will Novo know when he has become aware of his experiences' phenomenal properties as opposed to some other generic properties of his experience which are not phenomenal ones?

The realist may note that for all I have shown Novo *is* aware of his phenomenal properties, what he lacks is just awareness of them *as such*. But this does not help the realist. It is tantamount to saying Novo is aware of the target phenomenon, he just does not know he is (or he is just not aware that he is aware of it). This is an unhelpful possibility in the present context. It raises the question how the realist is justified in thinking the above is true of Novo. Was her training similar? If so, how did she go from not being aware of Good character as such to being aware of it as such? If there is an answer, she should be able to be share it with Novo. If there is no answer, the objection dissolves. (I further elaborate and defend the point in the context of defending premise 3 in section 5 later.)<sup>13</sup>

---

<sup>13</sup> This point applies to attention too. Suppose the realist wants to say that, really, all we ever attend to are phenomenal properties. That might be fine to say after it has been established that Good character exists. But if we are looking for a reason to believe they exist, it will not help to say we can attend to them in spite of its never seeming we are attending to them.

I contend that there is nothing Rea can add to her instructions to nudge Novo in the right direction. It will never seem to Novo that he has succeeded in becoming aware of phenomenal properties. (The obvious rejoinder to this contention is anticipated just below.)

This is to say, then, that even if we qualify transparency so as to accommodate the observation that we can become aware of more than external objects/properties in attending to our experiences, this will not avail Rea in her efforts to get Novo to discover what phenomenal properties are. If phenomenal properties are not the sorts of things to which subjects can attend, then it is plausible that, in the Good case, Rea will not be able to help Novo become aware of them.

Before continuing, I want to address the obvious rejoinder to what has so far been claimed. We can frame it rhetorically, as the following piece of incredulity: “Are you saying there’s nothing it’s like to perceive?!”

The following reply should suffice for the time being—but I will have more to say regarding the objection below in rounding out the case for premise 1. The reply is this: even if we grant that there is something it is like to perceive, if the novice cannot have his awareness guided to the what-it’s-likeness, and if he cannot become aware of it as such, then it will not matter that there is something it is like. In order to have a reason to believe Good character exists, the novice needs to be able to appreciate Good character as such.

### *3.1 Rounding out the case for premise 1*

With the above in place, I can begin rounding out the case for premise 1 of the argument against Good character. I will do this by establishing that there is nothing theory-neutral that Rea can say to aid Novo in discovering the phenomenal properties of his experience. Which is significant

given we are assuming the Independence Constraint—given we are assuming that whatever phenomenal character is, the road to it, so to speak, is nowhere paved in theory.

The goal of this section is to show that there is nothing Rea can say to aid Novo in becoming aware of Good character as such which respects the Independence Constraint. I will show this by considering the likeliest things the realist would say in their efforts to convey what Good character is. To be sure, there are indefinitely many things the realist might say in her efforts to give the novice a Good reason to believe Good character exists, and I cannot address them all. Instead, the plan for this section is address those things I take it that the realist is likeliest to say in response to the case so far towards premise 1. And to make the realist's job easier, let us work with a specific example: *what it's like to visually experience redness*. To cut down on verbiage, let us label the property 'phenomenal red'. The question for us is, What are the likeliest realist responses to the eliminativist's arguing that Rea cannot aid Novo in training his awareness on the phenomenal red of his experience?

I take it that the things the realist is likeliest to want to say in reply to what has so far been argued will cluster around the following three themes (and the things she might say that do not cluster around these themes will hopefully be handled in the Objections below):

- i) Seeing red is different from seeing blue, and this difference is a difference in phenomenal red/blue;<sup>14</sup> or

---

<sup>14</sup> This strategy was suggested to me by Will Davies (in conversation) and is reminiscent of how Tye begins his SEP entry on qualia (2019).

ii) phenomenal red is the directly recognizable or directly demonstrable feature of visual experience of red;<sup>15</sup> or

iii) phenomenal red is what Mary learns on first seeing red.<sup>16</sup>

(For prefatory purposes, Mary is a super-intelligent vision-scientist who has conducted all her scientific training without ever having seen chromatic colors for herself.) I will spend the next subsection paying special attention to (iii). I will focus just on it because enough was said in the previous sub-section to allow for brief handlings of (i) and (ii), which I turn to now.

Regarding (ii): the previous sub-section motivates taking our novice Novo to be unable to directly-demonstrate or directly-recognize phenomenal red as such. So far, he has succeeded just in directing his awareness to properties which before theory are not phenomenal. And if we think a condition on direct-demonstration/recognition is awareness, then Novo is not up to the task. Again, we might want to say he directly demonstrates/recognizes phenomenal red in spite of being unaware that he does (despite, moreover, thinking he is aware of something else). But this is irrelevant for the same reason it was irrelevant above that Novo may be aware of phenomenal properties but not as such. If Rea cannot get Novo to directly demonstrate/recognize phenomenal red as such, the question is raised how Rea ever managed the feat. (Again, more on this in section 5.)

And here is why we should not rely on (i) to convey to Novo what phenomenal red is. The salient *intentional differences* between seeing red and seeing blue will be too easily confused for what

---

<sup>15</sup> Cf. Loar, 2004; Chalmers, 2010, p. 256 ff for discussion of the direct demonstrability of phenomenal properties, and see Carruthers, 2000 for their direct recognizability.

<sup>16</sup> Versions of (ii) and (iii) can be found in Frankish (2012). See also Stoljar and Nagasawa (2004) for the idea that Jackson's Mary thought-experiment, even if not the best tool for arguing against physicalism, is still a good tool for conveying what phenomenal character is.



the realist is trying to allude to. The one experience is of/about red, and the other is of/about blue. And these intentional differences, before theory, are not phenomenal red/blue. This case against (i) is admittedly quick. But it will be more plausible in the light of what I will presently say regarding (iii).

### 3.1.1 *Mary*

Mary is an ideally intelligent color- and vision-scientist who acquires all her scientific learning in an achromatic lab without ever seeing a hued surface until one fateful day (Jackson, *op. cit.* p 130).

The idea is that Rea should tell Novo that phenomenal red is what Mary comes to know (is what Mary learns) on seeing red for the first time. In particular, she should proceed like so:

Before release from her achromatic lab, where she has spent all her life, Mary has never seen red. Mary is, again, of ideal intelligence and knows all there is to know regarding the sciences of color and perception. She knows that *these* are the spectral surface reflectances (SSRs) present, and *these* the viewing conditions, when subjects judge things red; *these* are the implicated neural states and *these* the relevant wide- and narrow-functional states; *these* are the optic and neurological laws; ... In short, she knows all relevant physical facts. On release from her lab, she sees a red thing for the first time, and in seeing the red thing she learns what it's like to see red. That is, in seeing a red thing for the first time, Mary comes to know phenomenal red.

Rea will tell Novo to put himself in Mary's shoes in the hopes that by his doing so he comes to understand what phenomenal red is. Once he understands what it is, he should be able to recognize it when it is present in his own experience. After this, he can use this strategy again to discover other phenomenal color properties.

### 3.1.2 *The rub*

To see why this strategy fails, start by recalling why the realist is evoking Mary here. In the previous section, the realist encountered a problem: the most natural strategy of conveying to the novice what (minimal) Good character is did not distinguish it from all sorts of distracting chaff. There were external qualities, intentional aspects of the experience, and affective responses to the scene perceived—none of which, before theory, are phenomenal properties. It was a problem that Rea could not winnow Good character from this chaff because Novo, being trained on the chaff, runs too great a risk of confusing Good character with those non-phenomenal properties. And if he does that, then when in future he discusses Good character with others, he runs too high a risk of talking past them. (Unless they have made his same confusions, but then they will not be talking about Good character.) And because Novo is not special, this will not be his fault.

So, the realist tells the Mary story to do a better job of guiding the novice to (minimal) Good character. To do a better job, the Mary story needs to make stark phenomenal red's difference from what is not phenomenal red. That way, the realist eliminates the risk of steering the novice to just more non-phenomenal properties.

With this in mind, we should be able to see why the Mary story does no better than the realist's first attempt (section 3) to convey what Good character is. Many realists (with the exceptions to be discussed shortly) accept that Mary cannot come to learn what phenomenal red is without learning *what redness itself is like*.<sup>17</sup> So, by these realists' lights, if Mary learns what phenomenal red

---

<sup>17</sup> It is worth pausing on this notion. By it, I mean something like *redness's own qualitative character*. But there is something slightly infelicitous about this latter expression. 'Character' just picks out agglomerations of qualities (quality-instances, really). But redness just is a quality. So, strictly speaking the quality will not *have* character, but will *contribute to* character. By 'what redness itself is like', I mean to refer to redness in its capacity to contribute to the making-up of character in this way; I mean redness qua quality. This is the role appealed to by naïve realists who say

is on seeing red for the first time, then she also learns what redness itself is like. (I will touch on what kind of learning this is—is it propositional? acquaintance-based?—below.) Redness, then, remains as distracting now as it was in the previous discussion, when Novo ran the risk of confusing the apple’s redness for Good character.

It might not be so bad if redness were all there were to distract the novice. But once we allow Mary learns what redness itself is like, she plausibly comes to learn a whole host of new non-phenomenal things.

To start, note that before release and before coming to know what redness itself is like, Mary will not have been able to think demonstratively about redness. After release, however, she will. And for being able to do that, she will know new propositions, ones like: “*this* is emotionally arousing”,<sup>18</sup> “*this* is brighter than black,” “*this* should go in my room,” and so on. Note moreover that before coming to know what redness itself is like, Mary would not have had what we might call a *color-character concept*, ‘R’, corresponding to redness’s being like what it is like, or to redness’s being *like such-and-such* (cf. Chalmers, 2010, p 258).<sup>19</sup> And for previously lacking such a concept,

---

that what redness itself is like ‘shapes the contours’ of Good character. And again, the realist who thinks no such thing exists will be considered below.

<sup>18</sup> Cf. Hardin, 1988, p. 167.

<sup>19</sup> Readers familiar with Chalmers’s (2010) taxonomy of *phenomenal* concepts will catch that I am simply re-appropriating his ‘pure phenomenal concepts’ to express the learning that for instance Byrne (2006), Tye (2009), or Johnston (2004), who each thinks Mary learns about redness, might like to attribute to Mary. Note well, however, that some realists may want to deny Mary acquires anything more than a demonstrative concept (or an indexical concept) on first seeing red (cf., e.g. Loar, 1997; Perry, 2001—discussed in Chalmers, 2010). The points I want to make can probably withstand this denial, but it is a denial the realist should not make. Mary’s new non-demonstrative color-character concept (‘R’) is the concept centrally implicated in her (new) power to *visualize* red things (among perhaps other conceptual powers). Cf. Williamson (2013), who argues the relevant science indicates

she would not have known propositions constituted by the content corresponding to that concept. But after learning what redness itself is like, Mary acquires ‘R’ and, so, can have attitudes constituted by it and think things like: “R is amazing;” “I want to cover a canvas with R;” and so on. Propositions like these and the above can be multiplied indefinitely. And, so, on seeing something red for the first time, Mary comes to learn *a lot*, which all comes to just more distracting chaff.<sup>20</sup>

---

that color concepts (of normally sighted subjects) underpin subjects’ capacities to visualize prototypical examples of the colours (295-96).

<sup>20</sup> An issue needs addressing regarding what kind of knowledge is at stake. There is a case to be made that Mary gains no new propositional knowledge, as I have had her doing, but only acquires new ways of knowing old things. It might, e.g., be argued that all that happens in the demonstrative case is Mary comes to be able to think with a demonstrative concept what before she could think with (say) a deferential concept. Contrasting new and old cases, we have: ‘[this] is emotionally arousing’ versus ‘[the redness with which my conspecifics are familiar] is emotionally arousing’. Suppose propositions are extensions, as on the Russellian and possible-worlds views. In that case, if the bracketed expressions co-refer, then the propositions known in each of the contrasted cases are plausibly identical, making it so that Mary learns no new propositions. (Of course, if propositions are intensions, as on the Fregean view, then this does not happen.) But the realist is saying Mary learns *something*. So, what Mary learns will have to be non-propositional (contra me). It could either be know-how that she learns, or it could be acquaintance-knowledge that she learns. The realist employing the present strategy (of conveying what phenomenal red is by telling the Mary story) will deny Mary’s learning consists just in know-how; after all, the know-how response to the Mary story was famously made to avoid saying she learns about phenomenal character (see, e.g., Nemirow, 1980; Lewis, 1988). So, to resist my response, the realist should say it is acquaintance-knowledge Mary learns. If her new knowledge is acquaintance-knowledge, then it is harder to make the points I am trying to make in the main text. (It remains just as easy, however, to make the point that she learns what redness itself is like.) But it can still be done. I would simply have to provide some maybe awkward-sounding paraphrases. For instance, instead of saying Mary learns the proposition *R is emotionally arousing*, I could say Mary becomes acquainted with R’s arousing-ness, or that Mary comes to acquaintance-know R’s arousing-ness. And so on, *mutatis mutandis*, with respect to the rest of the many alluded-to propositions. (Alternatively, I could say—sidestepping all the above fuss—that new ways of believing old things

Importantly, I mean for the new knowledge I am emphasizing to mark out salient aspects of Mary's learning, not banal aspects, like what she learns, for instance, by performing disjunctive addition ("R is amazing, or Trump beat Biden"). In emphasizing just what learned things would be salient for Mary on first seeing red, I mean things like how striking red is, how preferable to the nonchromatic colors it is, and so on.<sup>21</sup>

There is a final (salient) aspect of Mary's learning I want to emphasize, one that consists in her newly coming to have a distinctive kind of *understanding*. Consider a point familiar from mid-20<sup>th</sup> century Philosophy of Action, namely, that others' actions are not intelligible to us unless we understand their reasons for acting as they do.<sup>22</sup> (For a ready example, think of watching people play a game about which you know little to nothing. Very many moves will be unintelligible to you.) Now, suppose that what redness itself is like sometimes constitutes a *motivating reason* for various actions. Redness's being like what it is like constitutes a reason to paint the canvas *thus*; a reason to be irritated by the bedroom walls' being uniformly red; a reason to gift red roses; and so on and on. If what redness itself is like is sometimes a motivating reason for these actions, then the actions cannot be understood just on the basis of understanding their physical-dynamical explanations. But this is all that Mary had before release. Now that Mary knows what redness itself is like, however, she is in a position to appreciate the, as we might put it, rational

---

still count as learning. And, so, if Mary after release is capable of thinking about redness in new ways, then what Mary learns on first seeing red is still as much as was originally claimed in the main text.)

<sup>21</sup> Again, there is a sense in which she knew these things via testimony before release. But in that sense, Mary's knowledge was constituted by deferential concepts, or something thereabouts, distinct from the far richer concepts she now has at her disposal. If the points need to be re-framed to fit an acquaintance-knowledge conception of what Mary learns, I suggest how the paraphrases might go in the previous footnote.

<sup>22</sup> The points of this paragraph are influenced by Grimm (2016) who defends this idea in connection with ideas developed in Anscombe (1957). Cf. also Davidson (1963).

oomph of that character. She now understands and finds intelligible actions like the above that are rationalized by redness's being like what it is like. This is no minor piece of learning. It is important: Mary is coming to better understand her conspecifics' agency.

Above I said that Mary learns very much on seeing something red for the first time. Now it is hopefully clearer that she does. And, so, telling the novice just that phenomenal red is what Mary learns on seeing something red for the first time offers us no improvement to what was covered in the previous section. She learns an enormous amount of new and important things. The realist has failed to make stark phenomenal red's difference from all these important things.

As a reaction to the above, the realist might deny external redness sports its own what-it's-likeness. (The realists who do this are the exceptions alluded to above.) If Rea denies redness sports its own what-it's-likeness, she will presumably have a much easier time steering Novo to phenomenal red. Novo will no longer be distracted in the story by what redness itself is like because Rea is now saying it does not exist. And if what redness itself is like is no longer there to distract Novo, then neither will he be distracted by Mary's learning things constituted by what redness itself is like, like the learning that involved demonstrative thought, and like Mary's novel understanding of the agency of color-sighted subjects. Accordingly, if the realist denies redness sports its own what-it's-likeness, then presumably she will have uniquely singled out phenomenal red from any potential distractors (and in that way will have made stark how it differs from non-phenomenal properties).

What might entitle the realist to make the above denial? For one, she might be antecedently committed to *eliminativism-about-color*, the view that external color does not exist. Such a view might allow that the 'red' of everyday speech has a genuine referent (like a motley of SSRs, say), but it will deny the existence of external red-character. Alternatively, the realist might be motivated to make the above denial for holding a view on which color straddles the subject/non-subject divide but where color-*character* is strictly mental. As with eliminativism-

about-color, on such a view, the 'red' of everyday color-discourse might refer to an external property of objects, but 'color-character' will refer to something mental. The idea is that if Rea incorporates either of these views into her telling of the Mary story, then Novo will not be tempted to conflate phenomenal red with what redness itself is like (nor any of the knowledge noted above that is entrained by knowledge of what redness itself is like).

Still, relying on these views does not help the realist. And this because neither view, incorporated into the Mary story, points towards perceptual phenomenal character *minimally construed*. I will explain.

Consider the former view first. Eliminativism-about-color is usually paired with the postulation of *color qualia* in external color-character's stead (see Hardin, 1988; Kuehni, 1997; cf. also Chalmers, 2006). But if it is color-qualia that Mary learns about on seeing something red for the first time, then it is not just phenomenal red minimally construed that she learns about. This raises a few (related) issues. The first is this. Before theory, red-qualia are not minimal phenomenal red properties. Qualia, like SSRs or states that represent SSRs, are posited by phenomenal theorists to account for, or explain, minimal phenomenal red. Rea is trying to help Novo discover the phenomenon qualia are invoked to explain. She is trying to make stark how the thing to be explained differs from the things invoked to explain. Accordingly, leading the novice to red-qualia is to lead him to just more distracting chaff.

Second, recall that the Mary story is being evoked as a sort of ostensive definition of minimal perceptual phenomenal character (where what is being ostended is an aspect of a thought-experiment, sure, but one from which typical subjects should be able to extrapolate easily enough). Our ostensive definition for minimal perceptual phenomenal character cannot be most naturally taken to consist in *qualia's* being ostended. Otherwise, our task has (unwittingly) turned into an ostensive definition of qualia. Now, this same thing might have been said earlier, with respect to the Mary story's turning into an ostensive definition of external red-character. But the

problem feels even more acute here given the conceptual closeness of red-qualia on the one hand and minimal phenomenal red on the other. After all, qualia (we saw in section 1) are just phenomenal properties classically construed. Given as much, we should expect that, if anything, Novo will have an even harder time distinguishing minimal phenomenal red from red-qualia than he might have trying to distinguish external red-character from minimal phenomenal red.

And finally, minimal phenomenal properties were originally introduced by theorists antagonistic to qualia. The current strategy, then, is a bad one for realists generally. (This same point applies to the next view of color, too.)

For these reasons, incorporating eliminativism-about-color into the telling of the Mary story does not help Rea steer Novo any closer to minimal phenomenal red and, so, is not a helpful reaction to my saying Mary learns a lot on first seeing red.<sup>23</sup>

Similar considerations apply to views of color on which colors straddle the subject/non-subject divide, as on dispositionalism or relationalism. Take just the former view, which says that an object's redness (e.g.) is identical to the disposition of its surface to look red to normal viewers in normal viewing circumstances (Levin, 2000, p. 151). The view leaves open what looking red

---

<sup>23</sup> This view (and maybe also the next) fits only uncomfortably with Transparency, which I have assumed. On the view, red-character is an intrinsic property of Mary (qua subject). If Mary attends to redness on first seeing it, however, which it seems obvious she does, then redness is external and, so, not an intrinsic property of Mary. (Really, if *anyone* attends to color, then it is external.) Rather than appealing to Transparency to undermine the view, which might beg central questions, I want to cede that the realist can think of the assumption of Transparency as having all along served a largely heuristic role, one that allows us to bypass metaphysical thickets earlier on, but which can be dropped in the face of the present, inevitable thicket. Dropping Transparency earlier would have convoluted that discussion unnecessarily. Moreover, because most realists about minimal perceptual phenomenal character who are eliminativists about qualia (externalist representationalists; naïve realists) accept Transparency, assuming Transparency expedites the earlier discussion in a way that should not feel unfair.



amounts to so that it can be filled in as one's preferred theory of visual experience prescribes. But it does not leave things wide open. Looking red plausibly cannot consist in the caused experience's representing the object as red, or in the experience's involving the presentation of redness, on pain of circularity (cf. Levin, *ibid.* p 164; Peacocke, 1984, pp. 373, 375).<sup>24</sup> Looking red must consist in surfaces' causing experiences with a specific intrinsic qualitative property (Levin, *op. cit.*, p. 164; Peacocke, *op. cit.*).<sup>25</sup> Where this is what looking red consists in, what is plausible is that what Mary learns on first seeing red is what this non-representational, non-relational, intrinsic feature itself is like (analogously to her learning what a quale was like just above). Such a feature—a non-representational, non-relational, intrinsic mental quality—sounds much too like classical phenomenal properties for comfort. It seems that Mary is here too just learning about qualia. The realist's strategy again resembles more an ostensive definition of qualia than of minimal phenomenal red. We see a second time, then, that incorporating views of color like these into the Mary story does not help the realist.

(That these views lead us to qualia is their greatest sin. But they lead to Mary's learning a host of non-phenomenal things as well, things exactly analogous to the demonstrative-learning and the novel understanding discussed above.)

Let me recap. On seeing something red of the first time, Mary learns what redness itself is like (or what qualia are like, which does not affect the eliminativism on offer). In learning what redness itself is like, she learns a host of new things constituted by what redness itself is like. She

---

<sup>24</sup> Which alone is bad, given most realists accept one of these two perceptual theses.

<sup>25</sup> This, anyway, seems to be the orthodox understanding of how looking-red can be cashed out, compatibly with dispositionalism. But we might doubt orthodoxy here. Being red and looking-red can be mutually grounding if the mutual grounding is only partial. I do not want to pursue this point further, however, because the realist should not want to rest her case against premise 1 of my argument for eliminativism on the hope for such a dispositionalism working out.

also comes to be in possession of a new sort of understanding; she now understands those actions of the color-sighted that are motivated by redness's being like what it is like. Here is the main upshot of this. It is no good for Rea to evoke Jackson's Mary to steer Novo to phenomenal red. All of the distractions that scuppered the realist's first strategy (of guiding the novice to Good character by telling him he can train his awareness on it by attending to what is not it) find analogues in this new context. Before, the novice was distracted by external qualities and affective responses to those qualities. Now, he finds those again, plus more besides. So, if the earlier strategy failed, then the Mary strategy fails too.

### *3.3 Recap*

The case for premise 1 of our argument against Good character had a few stages. We were to put ourselves in the shoes of a newcomer to the Philosophy of Mind. This novice, as I have been calling him, asks a realist for help in conducting his attention to the phenomenal properties of his experience. The first stage of the argument appeals to perceptual experience's transparency to defend that the realist will be unable to give the novice a Good reason to believe Good character exists. Thanks to transparency, the realist cannot guide the novice's *attention* to his perceptual phenomenal properties because the only thing the novice will find he can attend to in introspecting his perceptual experiences are properties of the external objects of his distal environment.

The second stage of the argument draws out further difficulties which stem from transparency. For argument's sake, we granted that though we cannot attend to Good character, perhaps we can still become aware of it by attending to external properties. This is not helpful, I said, because there is very much of our experience and of our environments, which is not Good character, that we can become aware of by attending to the external properties of our distal environments. So, our novice will still yet to have been guided to Good character.

I finished the case for premise 1 by anticipating realist reactions to the case so far made for the premise. Those reactions could all be handled by considering how the realist might exploit Jackson's famous thought-experiment. This tack did not succeed, I showed, because the story of Mary does not winnow Good character from considerable distracting chaff.

We should accept, then, that Novo cannot be given a Good reason to believe Good character exists. And there is nothing unique about Novo. Novo represents all novices. Accordingly, we should accept that no novice can be given a reason to believe Good character exists.

#### **4. Premise 2**<sup>26</sup>

The second premise of the argument says that if no novice can be given a Good reason to believe Good character exists, then neither can he be given a Bad reason to believe it exists. There are indefinitely many Bad reasons the realist might come up with to believe Good character exists, and I cannot predict them all. So the goal is to look at those strategies the realist again seems likeliest to pursue. I will examine two types of strategy of providing a Bad reason to believe Good character exists. The first sees the realist providing an argument for Good character premised on i) the apparent obviousness of the existence of phenomenal character in the Bad case and ii) the Good case's indiscriminability from the Bad case. The second strategy instead stipulates what Good character is in relation to the phenomenal character intuitively had by the Bad case. Namely, that it is made up of those aspects of the Good case that are indiscriminable from Bad character. What I will argue in this section is this: treated as a deductive argument, the first strategy cannot avoid begging the question. Treated alternatively as an argument to the best explanation, the strategy makes Bad character overly partisan, and in that

---

<sup>26</sup> It is important to note that none of the strategies we will look at in this section are at all straightforwardly available to those realists who motivate their preferred phenomenal theses by appeal exclusively to considerations of the Good case, the way naïve realists do (cf. Sturgeon, 2008, pp.s 119-22). Because naïve realists make up a sizable sup-group of my opponents, I take this to bode badly for the strategies that follow.

way fails to respect the Independence Constraint. And the second strategy mis-identifies phenomenal properties in too many cases.

#### *4.1 Arguing for Good character*

It is not uncommon that Bad reasons for Good character tacitly invoke the following assumptions:

- 1) There is Bad character.
- 2) The Bad case is indiscriminable from the Good case.

My thinking this is not uncommon is influenced by a few arguments in the literature. First, Shoemaker (1994) provides an argument for the existence of phenomenal properties crucially premised on certain spectral inversion scenarios. Comparing the mental lives of spectral inverts gives us one argument for the existence of Good character. And because it is intuitive that the experience of at least one of a pair of inverts is Bad, Shoemaker's argument constitutes a Bad reason to believe Good character exists. Second, Loar (2002) develops a technique to get subjects around the obstacles to introspecting phenomenal properties generated by experience's transparency. If we can introspect the phenomenal properties of our experience (for the distinctive aspects of our experience they are alleged to be), then we will have resolved most of the problems I raised against realism in the previous section and will have found a reason to believe Good character exists. The key to introspecting Good character on Loar's account involves comparing the mental lives of brains-in-voids (BIVs) with our own. Because the experiences of BIVs are Bad, mastering Loar's technique would supply us with a Bad reason to believe Good character exists. And finally, arguments for sense-data, like the Argument from Hallucination (see, e.g., Robinson, 1994; see also the discussion in Crane and French, 2021, sect. 2.2), are arguments for classic phenomenal properties roughly premised on the above assumptions.

Each of the above arguments, however, are arguments for *internalist* theses of phenomenal character, where a phenomenal thesis is internalist only if it posits phenomenal character's constitution by exclusively intrinsic features of subjects. And arguably the lion's share of my opponents are *externalists*: they take phenomenal character to be partly grounded in external objects/properties. So, it will not help the realist about minimal Good character to rehash the above arguments. (But I will comment in section 5 on whether it would be worthwhile for the realist to posit partisan Bad character.) The details in any case should not detain us because the question for us is simply whether the above assumptions can in general be relied on by realists who wish to argue for Good character. I will start with whether the assumptions can be relied on in a deductive argument for Good character.

#### 4.1.1 Deducing Good character

The problem with any deductive argument for Good character premised on (1) and (2) is that the assumptions are not enough on their own to get us to Good character. To see this, note the point perhaps familiar from the disjunctivist literature on perception that indiscriminability does not automatically generate the needed link between Bad and Good character. On at least one important construal of indiscriminability (Williamson, 1991), two items' indiscriminability does not entail they have properties in common (excluding trivial ones, like *being mutually indiscriminable*). On Williamson's account, for  $x$  to be indiscriminable from the  $F$ s for a subject is for the subject to be unable to activate knowledge that  $x$  is numerically distinct from the  $F$ s (ch. 1; cf. Sturgeon, 2008, p. 127). Cashing out assumption (2) in these terms gives us the following assumption:

- 2\*) subjects are unable to activate knowledge that a Bad case as of  $\Phi$  is numerically distinct from the Good cases of  $\Phi$ .

But it does not follow from (1) and (2\*) that the Good and Bad case have non-trivial properties in common. In order for (1) and (2\*) to get us to Good character, the realist needs to

supplement the fact of the cases' indiscriminability. She would need to add, namely, that the reason subjects cannot activate knowledge that a Bad case is distinct from corresponding Good cases is *because subjects appreciate* a sharing of phenomenal properties across cases. But if this further assumption is what explains the Good and Bad cases' indiscriminability, then the argument which relies on it begs the question.

And the realist could not opt directly to cash out (2) in terms of the Good and Bad cases' sharing phenomenal properties, either. This would be plainly question-begging. Deductive arguments for Good character premised (1) and (2) seem bound to fail, then.

#### *4.1.2 Good character as the best explanation*

I turn to what at first blush is a more promising route to Good character. What is to stop the realist from taking the existence of Good character to be *the best explanation* of (1) and (2)?

The issue, I will argue, is that there is no way to motivate the existence of Bad character (assumption 1) such that the Good character which best explains (2), the indiscriminability of the Good and Bad case (hereinafter 'indiscriminability' for short), is a Good character sufficiently many realists will want to accept. And in this way, the argument to the best explanation will violate our Independence Constraint.

To see how the problem arises, suppose that the way we motivate (1) makes it seem we are meaning to suggest that the Bad case instantiates qualities which are strictly mind-dependent. In other words, that it instantiates qualities which do not consist in the familiar ones of external objects. (I consider an alternative below.) I think this is what would happen were we to invoke BIVs, hallucinations (as per arguments for sense-data), or inverts to motivate the existence of Bad character. But I say more on this below.

There are two initial problems with this way of motivating (1). First, too many realists reject that the Bad case involves qualities which do not at least partly consist in perceptible qualities of

external objects. I have in mind here representationalists, who hold that the Bad case is grounded in contents constituted by perceptible qualities. So, it would at least be unusual for these theorists to motivate the existence of Good character by appeal to these exotic properties. Moreover, these same theorists are inclined to accept that the Good case consists in subject's being aware of perceptible qualities of external objects. This raises the question why the *best* explanation for the Good case's indiscriminability from experiences which seemingly instantiate unfamiliar mind-dependent qualities is the existence of properties that, by most accounts, consist in totally familiar ones. If the best explanation is instead just more of what are strictly mind-dependent qualities of experience, then the inferred posit is not minimal Good character, but something non-minimal instead. And, so, it is incumbent on the realist who prefers this route to Good character to defend that the *minimal* posit really is the best explanation. I will return to this point after the following alternative.

What if we suppose instead that the way we motivate (1) makes it seem that the Bad case does involve perceptible qualities, but *uninstantiated ones*? Why uninstantiated? Because we could not make it seem we were meaning to suggest that the perceptible qualities involved were instantiated without making it seem we were describing a Good case rather than a Bad one. The question asked in the previous paragraph presses here too. Why is the *best* explanation of the Good case's indiscriminability from experiences which seemingly involve uninstantiated perceptible qualities the postulation of a character which by most accounts is constituted by perceptible quality instances? Best explanations should not immediately saddle us with mystery like this. And if the best explanation is instead just more uninstantiated properties, then the inferred posit is not minimal Good character, but again something non-minimal instead.

The realist might object that I have given no reason to expect that the way we motivate (1) should make it seem either that Bad character is mind-dependent (and apparently *sui generis*) or mind-independent (and uninstantiated). Why accept that our motivation is bound to point one

way or another? The simple answer is that this is just what happens in extant Bad reasons for Good character. Shoemaker's and Loar's arguments see us comparing our mental lives to those of subjects—certain inverts, BIVs—whose experiences could not involve familiar perceptible qualities. And in the Argument from Hallucination, we are to compare our mental lives to hallucinations, but where these are made to seem to involve what could not be familiar perceptible properties. Consider just the second premise of Robinson's (*op. cit.*) version of the argument: "The contents of hallucinations are *subjective images...*" (87, emphasis added).<sup>27</sup>

The onus, then, is on the realist to motivate (1) in such a way that she is not easily taken to be suggesting that Bad character is mind-dependent (and *sui generis*) or mind-independent (and uninstantiated). That way, the Good character that would best explain indiscriminability stands a chance at being minimal. Note well, however, that I have said enough already to motivate thinking this an onus the realist is unlikely to discharge. We need, after all, a feature unique to the Bad case which encourages thinking *its* phenomenal character will be easier to find than Good character. But the only intuitive unique such features, it seems, are *sui generis* mind-dependent qualities, or uninstantiated properties. And these, crucially, are posits invoked by phenomenal theses to accommodate Good character. So, to guide us to them is not to guide us to minimal Good character.

#### 4.2 *Stipulating what Good character is*

I turn now to the second strategy. Let us consider whether the realist could simply stipulate what the properties are which make up Good character. Specifically, could she stipulate that the properties which make up Good character are the ones that cannot be told apart from corresponding aspects of Bad character? Schematically, the strategy might look like this:

---

<sup>27</sup> See also the first premise of Crane and French's presentation of the argument: "An hallucinatory experience as of an ordinary object as *F* is not a case of awareness of an ordinary object" (*op. cit.*).



*Good phenomenal-F* is the aspect of Good experience of *F* that is indiscriminable from the aspect of Bad experience as of *F* it is in respect of which the experience is as of *F*.

To cut down on verbiage, let us label the aspect of Bad experience as of *F* it is in respect of which the experience is as of *F* '*F\**'. And we will assume that *F\** is a bit of Bad character. (If we do not assume this, then the stipulation strategy will look undermotivated for saying that Good character is what is indiscriminable from what may not be phenomenal character.)

The problem with this strategy is it seems to get the wrong results in paradigm cases. It *does* get the wrong results in the case that either of the mind-independent theses of color is true. (These are the views, touched on in section 3.1.2, at play in generating the verdict that Mary learns what external redness itself is like when she sees red for the first time.) To see this, let redness be our value of *F*. The current strategy then says that *Good phenomenal-red* is that aspect of Good experience of red that is indiscriminable from red\*. On the mind-independent characterizations of color, red objects sport their own mind-independent red-character. In the case that a mind-independent thesis of color is right, it plausible that *mind-independent redness* is that aspect of Good experience of redness which is indiscriminable from red\*. But I am assuming (section 1) mind-independent redness is not a phenomenal property. And, so, because we cannot rule out that a mind-independent thesis of colors is correct, neither can we ensure that the stipulation strategy correctly identifies the phenomenal properties of color experience.

It is not only colors with which the current proposal faces issues. If we treat the natural candidates for phenomenal properties as admissible values for *F*, then the schema gives bad results often. *Good phenomenal-square* is that aspect of experience of squares that is indiscriminable from squareness\*. But it is plausibly just perceptible squareness, that property of external objects, which is the aspect of Good experience of squares which is indiscriminable from squareness\*. This same thing goes with whatever perceptible shape properties we might take as our value of

*F*. Moreover, the same thing goes with whatever perceptible properties we might take as our value of *F*. The stipulation strategy, then, will not work.

A familiar point surfaces. There may be indefinitely many stipulation strategies the realist might pursue to convey what Good character is. But by showing the most natural such strategy to fail, I will have done enough to motivate thinking that the stipulation strategy is not how the realist should provide a Bad reason to believe Good character exists.<sup>28</sup>

The upshot, then, is that we cannot provide a Bad reason to believe Good character exists because, first, the most natural way of doing so, providing a Bad argument, is question-begging if deductive or, in the case that it is abductive, makes Bad character non-minimal. And the strategy of stipulating that Good character is what is indiscriminable from Bad character picks out, often, what may be purely external properties.

## 5. Premises 3, 4, and 5

I turn to the task of defending the remaining premises, 3 through 5. Premise 3, again, says

P3. If no novice can be given a Good or Bad reason to believe Good character exists, then no novice can be given a reason simpliciter to believe Good character exists.

---

<sup>28</sup> There is this further problem with the stipulation strategy, one I leave to a footnote because it affects only a smallish sub-group of my opponents, though an important one. The strategy cannot be pursued by those of my opponents who endorse the disjunctivism popularized in Martin (2004, 2006). (I continue to mean by '*F\**' the aspect of Bad experience it is in respect of which the experience is as of *F*.) On that view, *F\** consists in being indiscriminable from that aspect of Good experience of *F* it is in respect of which the experience is a Good experience of *F* (Martin, 2004, pp. 80-81). Conjoining this take on the nature of *F\** with the stipulation strategy, we are left with the following viciously circular take on what Good phenomenal-*F* is: it is that aspect of Good experience of *F* that is indiscriminable from indiscriminability from itself.

The premise is intuitive because if we can neither find a reason to believe Good character exists by considering Good or Bad cases, then our reason for believing Good character exists will have to come from reflection on things *other than* perceptual experience. But it would be extremely surprising if our reason to believe Good character exists consisted in reflection on things other than perceptual experience.

Perhaps, though, we think that realist testimony constitutes such a non-experiential reason. Why shouldn't the novice defer to the expert at the front of the classroom and simply accept the testimony of their realist instructors? If realist testimony is an admissible reason to believe Good character exists, then the question is raised what realists' reasons are for believing Good character exists. If their reasons are Good or Bad, then these reasons should be sharable. If their reasons are not Good or Bad, then their reasons might also consist in having previously accepted realist testimony. But a chain of reasons that consists purely in testimony makes the thing believed mythic (or worse, if myths are ultimately based in sincere but poor interpretations of events). And we ought not to add mythic items to our ontologies. Accordingly, we should require that novices' reasons for Good character ultimately bottom out in Good or Bad reasons. Without these, the novice is left without a reason simpliciter to believe Good character exists.

Premise 4 says that

P4. If no novice can be given a reason simpliciter to believe Good character exists, then none of us will ever have been able to be given a reason simpliciter to believe Good character exists.

Recall that by 'novice' I only mean a competent thinker yet to be familiarized with Philosophy of Mind. Once a person is familiarized with the subject, she is no longer a novice. So, the premise says that if we cannot be given, throughout the duration of our familiarization with the subject, a reason simpliciter to believe Good character exists, then those familiar with the subject—that is, *we*—also lack a reason simpliciter. I take this to be very plausible. But perhaps it will be objected that I have only shown that beginners, rather than novices, cannot be given a reason. Maybe we

think it is only after longish periods of training that anyone can become aware of Good character, and that I have not shown that longish periods of training cannot generate a reason.

This is not right. One way to interpret the above defenses of premises 1 and 2 is as my having showed that *there is no* viable training, or tutelage, that could result in a novice's having a reason to believe Good character exists. Maybe, for instance, we thought that that training consists in studying Loar's technique of "obliquely introspecting" (*op. cit.*) Good character by comparing our mental lives to those of BIVs. This, I showed, does not work. As with many of the strategies addressed above, there are conceivably indefinitely many trainings we could devise. But it suffices for my purposes just to consider those trainings which make use of the ingredients found in the strategies so far canvassed.

The final premise, 5, says that

P5. If none of us will ever have been able to be given a reason simpliciter to believe Good character exists, then none of us should believe Good character exists. (premise)

We need this premise because it does not follow from our having no reason to believe something exists that we should reject that it exists. For instance, you had no reason, before now, to believe that my friend E has a filling in a top left molar. But you certainly were not for that reason entitled to reject the proposition that she has that filling. There is a crucial difference in our case, however. Considerations of parsimony do prescribe rejecting the existence of types (or broad categories of types) that none of us has a reason to believe to exist. This is what I have argued is true of our circumstances with respect to Good character. None of us has a reason to believe to exist one broad category of types, namely, the various phenomenal properties thought to be instantiated by the Good case.

But let me emphasize that I am saying only that parsimony prescribes leaving out of our ontologies *types* that *none of us* (i.e., no one) has a reason to believe to exist. None of us has any reason to believe countless tokens of various species, for instance, to exist. But the parsimony

assumption I am making says nothing about whether those organisms should be left out of our ontologies. And many of us may lack reason to believe there is a species *Eurycea nana*. But some of us do. So the parsimony assumption I am making says nothing about whether that species should be left out of our ontologies.

Premises 1 through 5 entail that we should reject that Good character exists. The conclusion is perhaps an odd one. The most natural thing to expect from it is the practical implication that only Bad character exists. But the significance of the argument just presented discourages settling here. I will argue in this next section that concerns of motivation and parsimony encourage eliminating perceptual phenomenal character, minimally construed, altogether rather than settling for a view which posits Bad character alone.

## **5. Stage 2**

This brings us to the second stage of the case for eliminativism. Given we should reject that minimal Good character exists, what should we say about Bad character?

If we accept the upshot of sections 3-5, then the most the realist could say here is that though minimal Good character does not exist, Bad character does exist. This is an umbrella view, and it could be elaborated in a few ways. It could amount to i) a position on which Bad character alone exists, and it is minimal; ii) a position on which Bad character alone exists, and it is not minimal; and iii) a position on which minimal Good character does not exist, but non-minimal Good character does, and so does non-minimal Bad character. (I will not consider a position on which non-minimal, but not minimal, Good character exists, along with minimal, but not non-minimal, Bad character. It is a view in logical space but not one worth taking seriously.)

If the non-minimal character posited on the third view is something like qualia, then (iii) and its motivations fall outside the scope of concern of this paper. (Unless, that is, the motivation for the view comes strictly from the realist's being backed into a corner by the above arguments. In

that case, (iii) is exceedingly ad hoc and should not be adopted.) We are concerned only with phenomenal properties minimally construed, and qualia are not this. If the non-minimal character posited on (iii) is something besides qualia (like uninstantiated qualities, say), then the view, for never having been proposed before, is motivated strictly by the realist's being backed into a corner by the above arguments, in which case it is exceedingly ad hoc. So we can set (iii) aside.

How should we respond to the proposal of (i), a view on which just minimal Bad character exists but not also minimal Good character? Echoing a point made in section 4.1.2: there is apparently not anything unique to the Bad case which encourages thinking minimal character should be easier to find there than it was in the Good case. To see this, recall the realist strategies of conducting novices to Good character: novices were to become aware of it by attending to external properties, or by considering what Mary learns on seeing red for the first time. The natural Bad case analogue of these strategies would be to tell the novice he can become aware of minimal Bad character by attending to the qualities involved in the Bad case. But one of the lessons of section 4.1.2 was that there are no obvious non-tendentious qualities in the Bad case. So, our natural analogue, of telling the novice to train his awareness on Bad character by attending to the qualities of the Bad case, is not available. On top of this, there is no recognized attention-mechanism the inputs of which might be these qualities (non-minimal or not). This makes the view under consideration extremely hard to motivate. Accordingly, a view on which minimal Bad character alone exists is a poor option for the realist. There are a couple further problems for the view, but they are shared by the second option, (ii). So let us turn to that option.

How should we respond to the proposal of a view on which non-minimal Bad character alone exists? Crucially, a view on which non-minimal Bad character alone exists is not a view anyone has ever argued for. (This goes also for the first option.) So it is not independently motivated.

But perhaps it will nevertheless strike many as introspectively obvious that non-minimal Bad character exists—that the experiences of BIVs have it, that hallucinations have it, that certain illusions have it. Perhaps we think it is introspectively obvious that there are qualities unique to the Bad case in need of accommodation by theory. On having such a view put forward, we are faced with a choice. Do we embrace this view which has never before been argued for and that posits comparatively more property-types; *or* do we reject Bad character along with Good? The former view is imparsimonious with respect to the latter. (This goes also for the first option.) It is moreover not antecedently motivated, and it would attract few present-day realists. Going for a thoroughgoing eliminativism, on the other hand, has just the drawback of conflicting with what is *perhaps* introspectively obvious, namely, that there are qualities unique to the Bad case in need of accommodation by theory. Where these are our options, the choice is clear. The scale leans decidedly in favor of rejecting Bad character along with Good.

The realist's prospects if we accept the upshot of sections 3-5, then, are grim. Her remaining options are imparsimonious with respect to the alternative and are, more importantly, poorly motivated. This recommends doing away with minimal perceptual phenomenal character altogether.

## 6. Objections

### *Difference in what it's likeness between visual and tactile experience of shape*

Much of the reasoning involved in the defenses of premises 1 and 2 relies on the thought that it is too easy for novices to confuse what is supposed to be an aspect of their experience with what is a perceptible property of the external scene. This first objection starts by noting that this is all probably a symptom of the examples' being experiences of perceptible properties detectable by just one sense modality. Novices would be much less tempted to make these confusions were they to consider experiences of perceptible properties detectable by more than one modality. And, so, consider shape perception. What we should tell the novice is to consider the difference

between, for example, visually experiencing something square, and tactilely experiencing something square. Visual experience of square and tactile experience of square do not differ in respect of what they are of or about, namely, squareness. They can only differ in respect of phenomenal character. If we tell the novice this, then we will have given him a reason to believe Good character exists.

*Reply*

In response, I think we should follow Thomas Reid, who in his (1764) wrote:

“[The] blind man’s notion of visible figure will not be associated with colour, of which he hath no conception; but it will perhaps be associated with hardness or smoothness, with which he is acquainted by touch. These different associations are apt to impose upon us, and to make things seem different, which in reality are the same” (Reid, 1764/1997, Chapter 6, Section 7, quoted in Green, 2020, p. 18).<sup>29</sup>

The idea is that the differences in experience of some single shape property ultimately boil down to what are irrelevant, distracting differences. To the extent that subjects are compelled to judge visual and tactile experiences of shape different, this has just to do with visual experience of shape being inextricably bound up with experience of color, contour, shading, and depth, *inter alia*—and with tactile experience of shape being inextricably bound up with experience of smoothness, pressure, temperature, *inter alia*. What is important for our purposes is that my opponent needs Reid’s insight to be definitively false. Otherwise, she will not have given us a reason to believe Good character exists. And showing the insight to be definitively false strikes me as practically impossible. There after all have only ever been counter-intuitions waged against

---

<sup>29</sup> This insight is echoed in a recent paper by O’Callaghan (2019):

it is not obvious that the visual experience of sphericity—when it is considered in abstraction from the experience of other visible features—must differ in respect of each phenomenal feature tactual experience of sphericity“ (126, quoted in Green, 2021).



it (Block, 1996; O’Dea, 2006—both cited in Green, 2020; and Lopes, 2000, p. 445—cited in Green, 2021) but no arguments.

What’s more, in his (2021), E.J. Green argues that at early levels of visual and tactile (what he calls “haptic”) processing, visual experience and tactile experience of shape are in the business of putting subjects in contact with distinct shape-properties. At the earliest level of processing, visual experience, he argues, is experience of *solid-angles* (2021, p. 6), which one gets by taking the rays that emanate from a subject’s viewpoint which tangentially graze the perceived objects’ outer boundaries; the solid angle is the angle whose vertex is the viewpoint. And tactile experience of shape is experience of *somatotopic projections*, of, that is, pressure applied (directly or indirectly, as when something is felt by the use of a device, like a cane) by the object to body parts  $p_1, \dots, p_n$  (*ibid.*, p. 8). If this is right, then visual and tactile experience of shape differ in respect of what they are of/about. We would accordingly not have a case of distinct experiences’ being of or about a single property which experiences introspectively strike the subject as being different.

#### *Differences in focally attending and peripherally attending, and blur*

The next objection says that it should be easy to direct the novice to the difference between focally attending to an external property and non-focally attending to it, as when we are aware of something in our periphery. Take, for example, a case of seeing a dog directly before you and a case of being aware of a dog in your periphery, without focally attending to it. These experiences differ in what they are like, even though they are of or about the same property. Accordingly, if we can get our novice to become aware of this difference, then he will not confuse these aspects of his experience for what the experiences are of, and we will have guided him to Good character.

#### *Reply*

Seeing focally versus peripherally *is* different. It is epistemically different: peripheral awareness affords substantially less knowledge of our environments than focal awareness (perceptual attention) and is informationally much poorer. Similar differences crop up in seeing things far away versus close up or seeing things blurrily versus clearly. Peripherally attending is also different from focal awareness. It too is different epistemically, though less markedly. Other differences between focal and peripheral attention include the latter's involving noticeable strain of the muscles in the eye and the latter's being maybe a little disorienting.

Arguably, the 'what-it's-like' locution employed in the objection picks up on subjects' being able to know, reflecting just on epistemic contexts, about the differences between peripheral and focal experiences. But if the differences are just the ones noted above (epistemic, muscular, equilibrioceptive) then, because they are resources theorists call upon to accommodate phenomenal character, the objector needs to say more to motivate positing phenomenal character in addition.

The above notwithstanding, cases of visual blur and the like strike me as a thin reed on which to rest too much of consequence. Even supposing the current objection successful, it would remain true that the novice could not be made aware of any paradigm phenomenal properties, like the ones associated with visual experience of color, shape, and motion, or auditory experience of pitch, volume, and timbre. And blur phenomenology does not at all straightforwardly ramify for the existence of those. Accepting blur phenomenology probably does prescribe reducing one's confidence that paradigm phenomenal properties (minimally construed) do not exist. But it is far from obvious that accepting blur phenomenology prescribes accepting that paradigm phenomenology exists.<sup>30</sup>

*Does my line of reasoning preclude awareness of experience's intentional aspects?*

---

<sup>30</sup> For helpful work on the relationship between confidence and acceptance, see Kaplan (1981).

Above, I said that subjects could not be made aware of phenomenal properties because they will just be made aware of external properties and, I also granted, experience's intentional aspects. The present objection asks if my line of reasoning over-generates. Shouldn't subjects be unable to discover the intentional aspects of their experiences for the same reason they couldn't discover its phenomenal aspects?

The objection cuts in two ways. It works as a *reductio* by suggesting that any account which precludes subjects' being aware of intentional aspects of experience is too implausible. And it works by kicking out rungs of the ladder I used to reach eliminativism, because subjects' being potentially drawn to the intentional aspects of their experiences was part of what made it difficult to steer their awareness to phenomenal properties.

#### *Reply*

The reply here mimics Tye (2014), noted above. What my line of reasoning may do is preclude subjects' being made *de re* aware of experience's intentional aspects. But it does not preclude subjects' being aware *that* their experiences are of, or about, what they are of/about. And there are many accounts of how subjects attain knowledge *that* their experiences are of, or about, such-and-such, and it is not incumbent on me to supply any specific one. The eliminativist is free to pick from among them as suits her purposes. Moreover, there is no reason to suspect that the outputs of the awareness we have of the intentional aspects of our experience should not be distracting, as I relied on them as being, just for this awareness's failing to be *de re*.

#### **Conclusion**

The verdict, then, is that we should eliminate perceptual phenomenal character, minimally construed, altogether. None of us has a reason to believe that Good character exists which reason is meant to point to something that exists independently of theory. And considerations of parsimony and motivation prescribe rejecting Bad character too.

Because the conclusion is bound to be controversial, I want to mention here some things I mean not to be entrained by it. First, I have not said that there is no reading of ‘what-it’s-like’ that applies truly to experience. Where ‘I know what  $x$  is like’ is made true by knowing various of  $x$ ’s properties, or by knowing that  $x$  has various properties, we can say that subjects know what experience is like when they know various (non-phenomenal) properties of experience, like the properties that make up its epistemic profile; or when they know that various properties are typically instantiated by experience, again, like the ones that make up its epistemic profile, or like the ones that make up its intentional profile. (That I am committed to this much was intimated in my reply to the objection from blur.)

Second, to repeat a point made in the introduction, I did not argue for the elimination of minimal phenomenal character that is not perceptual. I did not argue that phenomenal character minimally construed cannot be made sense of in the case of pain, or ennui, or what have you. Given what all was said above, the reader might be able to predict how such arguments might go. But it is not a part of the current project to extend the argument in that direction.

And maybe most importantly, I do not take my argument to have particularly damning consequences for the major parts of the major perceptual theses. Even if minimal perceptual phenomenal character does not exist, there is still an important debate to be had regarding how to characterize the Good case. Is it fundamentally relational, or representational, or something else still? Moreover, even if minimal perceptual phenomenal character does not exist, there are still important questions to ask regarding hallucination: does it fundamentally consist in indiscriminability (from Goodness), or representation, or something else? And perceptual illusions: first off, are there any (see Kalderon, 2011, discussed in the following chapter of this thesis, for the answer that there are none)? If there are any, are the illusory experiences relational, representational, or... I develop this idea further in the following Afterword.

## *Afterword*

In the thesis's introduction, I briefly commented on what eliminativism prescribes not doing. It prescribes not investing further research into the grounds of phenomenal character; and it prescribes not investing research into what are phenomenal character's neural correlates. But I want to dedicate this post-script to saying something positive about how I think the argument of this paper ramifies for debates on perceptual consciousness.

A consequence of the paper is that there is no phenomenal character distinctive of, for example, visual experience of the color green. But it is manifest to experience that greenness exists.<sup>31</sup> Here, then, is a question that could, *partly*, replace the question of what grounds the phenomenal character distinctive of such experience:

Q1) What are we aware of when we are aware of greenness?

The usual candidate answers are as live as ever: we are aware of a surface spectral reflectance (SSR); or a primitive color property; or a disposition to bring about the instantiation by experience of a green-quale; or a green quale, among some other answers. (These views were discussed in section 3.2 of this paper.) In addition to this question, there remains the question of what perceptual awareness consists in.

Q2) When we are aware of whatever green is (SSR? quale?), in what does this awareness consist?

Is the awareness representational, fundamentally consisting in the having of satisfaction conditions (as we get in Dretske, 1995; Tye, 1996; Byrne and Hilbert, 2003)? Or is it instead

---

<sup>31</sup> In case disambiguation is called for, following Pautz (2020) we can call the greenness of which human subjects, in experience, are manifestly aware 'sensible greenness'. Pautz says this is what Byrne and Hilbert (2003) and Tye (2000) would call "the color green" or just "greenness", but what Shoemaker (1994) would call "phenomenal green", and Chalmers (2010) would call "perfect greenness".

relational, fundamentally consisting in subject's being related to greenness by way of an unanalyzable relation of perceptual-acquaintance (as we get in Martin, 2002; Brewer, 2011; Soteriou, 2013)? Or something else?

There are many combinations of answers to Q1 and Q2 which together replicate a good portion of the familiar geography of answers to the question of what the nature of perceptual phenomenal character is. For instance, there is a view on which the phenomenal character distinctive of visual experience as of green consists in subject's being unanalyzably related to a primitive green property. And compatibly with the eliminativism here on offer, there is a combo-view on which greenness, the property we are aware of in ordinary color-experience, is a primitive color property and on which that awareness consists in an unanalyzable perceptual relation, involving the subject's being acquainted with the primitive color property. To give another example, there is a view on which the phenomenal character distinctive of visual experience of green consists in subject's representing (what we can call) the green-SSR, where this representation consists in the having of contents, where contents are construed as sets of possible worlds. And compatibly with the eliminativism here on offer, there is a combo-view on which greenness, the property we are aware of in ordinary color-experience, is a SSR and on which awareness of the green-SSR consists in the successful representation of the green-SSR, with representation again construed as the having of contents that are sets of possible worlds (specifically, the set of worlds in which the experience accurately portrays things).

Let's call the geography of combined answers to the above Q1 and Q2—what are colors? and in what does awareness of them consist?—the *eliminativist landscape*. And let's call the current geography of answers to the question of what phenomenal character consists in the *realist landscape*. Compatible with the eliminativism on offer is the replication of *a piece* of the full realist landscape. Crucially, however, the eliminativist landscape does not replicate *all* of the realist landscape. If it did, then the preceding paper would amount to a rallying-cry to provide longer (if

more accurate) formulations of existing positions of the realist landscape. Basically, the paper would be a lot of fuss over something minor.

We can begin to see how the eliminativist landscape differs from the realist landscape by considering the family of views on which perceivers represent the external world thanks to phenomenal properties their experiences instantiate. Views that fall into this family of views include those on which phenomenal properties are essentially representational (for a survey of views of this sort, see Kriegel, 2013), and those on which phenomenal properties have their representational properties contingently (Papineau, 2021; Block, 1996). At first blush, motivating views like these after eliminativism is embraced becomes a probably too-delicate matter. We can see this by asking Q1 and Q2. To the question What are colors? (Q1), any answer could work. Let's suppose colors are something external (like SSRs). But once we ask the follow-up question, In what does awareness of color consist? (Q2), things quickly fall apart. If colors are external, then awareness of them presumably consists in their being represented by qualia, given the view under consideration. But what are qualia? What have they been invoked to explain? Presumably, they have been invoked to explain phenomenal character. But there is no phenomenal character in need of explaining, given we are embracing eliminativism. So, we lose the motivation for answering Q2 in that way. And we have begun to see some of the difficulties one might encounter in motivating a combination of answers to Q1 and Q2 that would replicate a thesis belonging to the above family of views.

Consider now that the answer to the question What are colors? (Q1) is "something internal" (like states of V4). If colors are internal, then it would be quite strange to answer Q2 (In what does awareness of color consist?) by appeal to representational qualia. No one has ever floated that qualia represent internal colors. What is more common is that internal colors are simply identified with qualia (as we get, for instance, in Kuehni, 2003). But even if a proponent of the above family of views were happy to float such a combination of views, she would run into the

problem noted just above: Why are qualia being invoked? Presumably, they are being invoked to account for phenomenal character. But there is no phenomenal character to explain. It is probably better just to identify colors with qualia. But then awareness of them will (presumably) not consist in qualia representing colors, which are here just further qualia.

And so we see again that it is difficult to arrive at a combination of views which would replicate a thesis belonging to the family of views under consideration. But these are major views in the realist landscape. So, the result that they are stripped of much of their motivations is a huge consequence of the eliminativism on offer. And it is why we should not take the eliminativism to be little more than a call for linguistic readjustments of the realist landscape. In case it needs clarifying, I am not saying it is difficult to motivate theses which posit qualia. On the contrary, such theses are easy to motivate. If colors are mental, they are plausibly just qualia. So motivating qualia should be no more difficult than motivating the identification of colors with mental states. Alternatively, if colors are dispositions to bring about experiences which instantiate qualia, as on a Lockean secondary-quality thesis of color, then positing qualia should be no more difficult than positing the identification of colors with dispositions to bring about experiences that instantiate qualia. What the above does show, then, is not that it is difficult to motivate positing qualia if eliminativism about perceptual phenomenal character is embraced. It shows, instead, that it is difficult to motivate the positing of representational qualia. And views which posit representational qualia mark out a large sub-region of the contemporary realist landscape.

The proponent of representational qualia might complain that my Q2 is unfair. She might respond instead as follows: “We are not *aware* of primitive color properties. Instead, we represent the external world as instantiating primitive color properties. Perceptual *contact*—but not perceptual awareness, which has a ring of factivity about it—with external reality consists in representing of it that it instantiates primitive color properties.”



That sounds like something that can still be said, even where eliminativism is embraced. But it shirks a question that demands answering. We just *are* aware of sensible greenness. So we need to have an answer as to what that awareness consists in. The proponent of representational qualia might want to respond as she has. But even she will have to acknowledge a need to respond to Q2. This because there must be some introspective reflection that yields the verdict that the world could not instantiate that primitive color properties we represent it as having. There must be a mechanism that allows us to introspect the qualities. Otherwise, our confidence that the world could not instantiate *this stuff*—this color-character—is totally misplaced.

To be sure, there is much more to be said, and the above ideas require elaboration. But because the main goal of the present work was just to argue for eliminativism, I will leave that cartographic project for another time.

## Disjunctivism and supersaturated red

**Abstract:** Experience as of supersaturated red, an afterimage-experience as of a red surface that is beyond-maximally saturated, presents a problem that is yet to have been successfully discharged by proponents of reflective disjunctivism, the view that perceptual experiences are to be fundamentally characterized in terms of their indiscriminability from perceptions. The experience presents a challenge for the view because there are no perceptions of hues that are beyond-maximally saturated from which might ground the experience. There being no perceptions of supersaturated surfaces also means that the naïve realist resources which constitute the backbone of reflective disjunctivism will not be available to account for the experience either.

Naïve realism grounds the so-called Good case of perceptual experience in subjects' standing in irreducible perceptual relations to aspects of mind-independent reality. What it's like for a subject to see a bright red apple on a matte, brown wooden table, for instance, is grounded in the subject's standing in an irreducible visual relation to the apple, the table, and their mind-independent visible properties. But the thesis cannot say this same sort of thing about the perceptual Bad case, of hallucinations and certain illusions. When a subject hallucinates, e.g., a pink elephant, there will often enough be no candidate aspects of mind-independent reality relation to which might ground what the hallucinatory experience is like. And this generalizes. So, the philosopher committed to the Good case's being grounded as per naïve realism will have to say something else about what grounds the Bad case.

Enter the winningly economical reflective disjunctivism (prominent supporters of which include Martin 2004, 2006; Fish, 2008, 2009; Brewer, 2011; Soteriou, 2013). Reflective disjunctivism maintains that the Good case is grounded as per naïve realism, in subject's standing in irreducible perceptual relations to aspects of mind-independent reality. But it adds that the perceptual Bad case, like the above hallucination, consists in nothing more than indiscriminability from Goodness. Significantly, indiscriminability is typically given an epistemic gloss whereunder some things are indiscriminable exactly if subjects in certain idealized circumstances cannot know the things to be numerically distinct. A hallucinatory experience as of a pink elephant on the view consists in the experience's being indiscriminable from the Good cases of seeing a pink elephant. Grounding the Bad case like this—in certain inabilities-to-know—makes the view economical. It obviates having to ascribe to subjects any “positive” states of knowledge as being involved in subjects' capacities to correctly classify Bad experiences as experiences (there will be much more to say about this in section 1.1).

Experience as of supersaturated red presents a problem for disjunctivism, one that disjunctivists have yet to successfully discharge. The experience is an afterimage-experience as of a red surface

which exhibits a level of saturation that is beyond maximum. It is one type of a class of afterimage-experiences each of which is as of a hue that is beyond-maximally saturated. These experiences present a challenge for disjunctivism because beyond-maximal saturation (“supersaturation”) is a metaphysically impossible property and, so, there are no Good cases of perceiving hues that are beyond-maximally saturated. But reflective disjunctivism grounds Bad experience in its being indiscriminable from such Good cases. The view, accordingly, runs into trouble in assigning these experiences the appropriate character. Experience as of supersaturated hue thus prima facie serves as a counterexample to the theory. But I will be restricting my attention just to supersaturated red.

I will be making the case that disjunctivists have yet to successfully accommodate experience as of supersaturated red. This will involve taking a close look at the two extant attempts in the literature to do so. The first, due to Martin (2004), appeals to resources unique to disjunctivism to accommodate experience as of supersaturated red. Martin says we should think of experience as of supersaturated red as decomposing into constituent elements each of which is not problematic from the perspective of disjunctivism. The next attempt to accommodate the experience, due to Kalderon (2011), appeals to resources of disjunctivism that it shares with naïve realism, which theory’s commitments comprise half of disjunctivism’s. Kalderon argues we should ground experience as of supersaturated red in subjects’ being related to afterimages’ objectively looking supersaturated red. Ways that things look, for Kalderon (and others: Brewer, 2011; Martin, 2011; Travis, 2004), are objective features of mind-independent reality. And, so, the proverbial stick half-submerged in water objectively looks bent, where its looking bent is a feature it possesses independently of the actual existence of minds. And if experience as of supersaturated red is grounded in subjects’ relations to these looks, then disjunctivism is not threatened by experience as of supersaturated red.

Crucially, the extensional inadequacy of reflective disjunctivism I mean to bring out is not paltry. As we will see, central to the view's motivation (section 1.1) is the presumption that the view is extensionally adequate (and that, really, it is rival views which run great risks of extensional adequacy). This makes it so that a failure of extensional adequacy sends a shockwave through the very foundations of the view, thus toppling the edifice on top.

The plan for the paper, then, is as follows. In the first section, we will get clear on what motivates acceptance of naïve realism and, subsequently, reflective disjunctivism. In section 2, experience as of supersaturated red is introduced, and the contention is made that it *prima facie* serves as a counterexample to the theory. And in the following section, 3, I will explain why appearances here are not deceiving: experiences as of the metaphysically impossible do indeed serve as counterexamples to the theory. In section 4, I will introduce the two extant attempts to respond to the challenge these experiences pose for reflective disjunctivism before arguing that neither response is successful. In the final section, I explain why experience as of supersaturated red is more than a mere counterexample for reflective disjunctivism. The problem the experience poses for the view undermines its very motivations, thus giving a double-edge to the problem.

### *1. Reflective disjunctivism*

Before I introduce naïve realism, which theory constitutes the backbone of our target theory, a pair of terms needs introducing. Following Sturgeon (2008), we will say that an experience is one of the *Good cases* to the extent that its conscious presentational aspects are determined by perceptual contact with the world (cf. p 113). And, so, take again the above visual experience of the apple on the table. This experience is Good to the extent that what it's like for the subject to have the experience is determined by successful visual contact made with the apple and table and their visible property-instances (hereinafter I will drop "instances"). If this is how we cash out Goodness, then we should say that an experience is one of the *Bad cases* to the extent that its conscious presentational aspects are not fixed by such contact with mind-independent reality.

(*N.b.*, I do not assume Bad cases so defined exist.<sup>32</sup> But it is important to leave them on the table for now because appealing to Bad cases later gives the disjunctivist more resources with which to accommodate the problem experience.) Supposing Bad cases so defined exist, hallucinations are Bad (Fish, 2009; Brewer, 2011; Soteriou, 2013). And perhaps so are certain illusions (Fish, 2009).

Now, one cannot get to the theory we will be principally concerned with, reflective disjunctivism, but by way of naïve realism. So, first: *naïve realism* (NR) for our purposes says that

**NR** the phenomenal character of the Good case fully consists in subjects' standing in irreducible perceptual relations to mind-independent objects and/or their properties, indexed to perceptual circumstances.

Significantly, the relation invoked in NR is not to be analyzed in terms of representation (cf. Beck, 2019, pp. 608-9; French, 2014, p. 395). The perceptual relation which grounds phenomenal character on NR may ground an experience's representing the parts of the world that the experience is a relation to. But the relation itself is not in the first instance representational. This is part of the relevant relation's irreducibility. For the relation to be irreducible is for it not to be grounded in, or decomposable into, further psychological or cognitive phenomena (Sturgeon, 2000, p. 10; French, 2020, p. 108). And the relevant notion of representationality is one such further phenomenon.

To illustrate NR, take, for example, a Good visual experience of a ripe Red Delicious perched on a matte, brown wooden table. NR says that what it's like for the subject to see the apple on the table is nothing over and above her standing in an irreducible visual relation to the apple and the table. Features of the apple, the table, the subject, and the visual relatedness of the subject to the objects: these are all constituents of the experience's phenomenal character.

---

<sup>32</sup> See Raleigh, 2014; Ali, 2018; Masrour, 2020, for arguments that, so defined, Bad cases do not exist.

A final feature of NR warrants mention, one relevant to the discussion in section 4.2. If perceptual experience of objects and their properties consisted *just* in perceptual relations to those objects and properties, then the worry might arise that important features of perceptual experience get left out of our account. For instance, the following aspects of perceptual phenomenology would be left out: it might seem to subjects who are visually perceiving a coin-shaped object from the side that they are seeing an ellipse rather than a circle; similarly, it might seem to subjects who are visually perceiving a square shaped object from above that they are seeing a rectangular surface and not a square one; and so on and on. If we want our account of perception to account for phenomenological aspects of perceptual experience like these, then we have to index our irreducible perceptual relations to *perceptual standpoints* (see Beck, *op. cit.*, pp. 609-10 for a survey of the many proponents of NR who invoke standpoints to do this). A perceptual standpoint will involve things like the subject's orientation with respect to the perceived object, like lighting conditions in the visual case, and like conditions of the transmitting medium in the auditory case, inter alia. I adhere to custom here in characterizing NR similarly.

A handful of considerations motivate NR. First, there are phenomenological motivations: introspective reflection on the transparency of experience delivers the verdict that bits of mind-independent reality shape the contours of what that experience is like for one.<sup>33</sup>

Second, there are epistemological motivations. Naïve realism is the only phenomenal thesis with the resources to illuminate the epistemic contact subjects have with worldly character. And for that, it is to be preferred to rival phenomenal theses, which either obfuscate this epistemological contact or otherwise leave it obscure.<sup>34</sup>

---

<sup>33</sup> Martin, 2002; cf. also Soteriou, 2013, who motivates NR by appeal to the transparency of temporal experience.

<sup>34</sup> For motivations in this vein, see Cambell, 2002; Brewer, 2011.

And third, there are semantic motivations. Naïve realism is the only phenomenal thesis with the resources to illuminate the semantic contact subjects have with worldly character. It illuminates subjects' abilities, particularly, to demonstrate worldly character in thought. And for that, it is to be preferred to rival phenomenal theses which either obfuscate this semantic contact or otherwise leave it obscure.<sup>35</sup>

Crucially, each of these motivations is world-directed. None of them consists in reflection on contributions made by the subject to what her experience is like. Each consists just in reflection on how specifically *worldly* character is presented, known, or thought about. Neither moreover does any of them consist in reflection on what illusions or hallucinations are like. Each consists just in reflection on the Good case.

There are different ways we could further flesh out NR (see, e.g., Beck, *op. cit.*, pp. 608-11 for discussion), but it is just one further development of NR that will occupy us. Before we get to it, note that a failure of NR to accommodate experience as of supersaturated red would just recommend calling the experience Bad. This because NR makes no mystery of the Bad case's falling outside its remit. It is transparently not built to handle such cases. So a failure of NR to accommodate a specific Bad case would not count against the theory (not, at least, without much further argument). Before the turn of the millennium, the historically most popular way to remain committed to NR whilst recognizing the need to account for the Bad case was to plump for what I will call *run-of-the-mill disjunctivism*. Phenomenal theses in this vein are minimally committed to the following two theses: i) the phenomenal character of the Good case is grounded as per NR, but ii) the phenomenal character of the Bad case is not. Anything could satisfy (ii) in keeping with run-of-the-mill disjunctivism, but some of the historically proposed candidate grounds have been: unworldly objects' appearing to subjects (Alston, 1999; Langsam,

---

<sup>35</sup> For motivations in this vein, see Putnam, 1999; Cambell, *op. cit.*; cf. also Brewer, 2011, ch. 3, where the point is made that indirect realist theses obfuscate the semantic contact subjects have with worldly character.



1997), unworldly facts' appearing to subjects (a view Sturgeon, 2008, attributes to McDowell), or neural features of subjects (Logue, 2013). But (ii) could alternatively conceivably be satisfied by functional role, extensional content, intensional contents, and so on.

We see, then, see that there are many and various ways of being a naïve realist. But among these ways, it is just with *reflective disjunctivism* that we will be concerned throughout (I motivate this restricting of attention below.) Reflective disjunctivism conceives of perceptual experience as follows:

**RD**      to be a perceptual experience as of  $\Phi$  is to be indiscriminable from the Good cases of  $\Phi$ .

with the Good case here understood to be grounded as per NR. One of the naïve realist responses to experience as of supersaturated red is conducted in terms of RD, so let us for a moment turn to why some naïve realists commit to the view.

RD is motivated by a certain modesty. The motivation starts by contrasting how two conceptions, one minimal and one not, answer an epistemological question regarding the having of perfectly matching hallucinations. On the minimal conception, which is just RD, to be a perceptual experience is to be indiscriminable from a Good case. Now consider this non-minimal foil: to be an experience is to exhibit experience-making features  $E_1, \dots, E_n$ , where these might be qualia, representational properties, or any of the rest of the usual suspects. As it turns out, the minimal conception is modest with respect to its non-minimal foil in answering an epistemological question about the presence of perceptual experience.

We get to this question by considering a lesson of Descartes, namely, that right now you might not be enjoying a Good case of the scene before you but instead be having a very vivid dream as if of that same scene. Such a case would be (at least for our purposes) a perfectly matching

hallucination (Martin, 2004, p. 47). With this Cartesian datum in mind, here is the question. How is it you know you might be victim to a perfectly matching hallucinating?<sup>36</sup>

On the minimal conception, the answer is: because I cannot tell I am not enjoying a Good case. Whereas on the non-minimal conception, the answer—this time two-part—is: because I can detect the presence of experience-making features, *and* because I know those features can be instantiated independently of Good cases. The minimal conception gives the comparatively modest answer. All that is required to know I might be hallucinating is a failure to know something, namely, that I am not perceiving. No positive states of knowledge are called for. Whereas on the non-minimal conception, I first need to know when certain features are instantiated, and, second, I need to know something about the *nature* of those features—about their modal profile—namely, that they can be instantiated independently of Good cases. This is more in the way of epistemic powers required of subjects in their knowing it is possible that they might be hallucinating. A tick, then, against the non-minimal conception of experience.

The proponent of the non-minimal conception might wonder whether the minimal conception really is extensionally adequate, which it must be if it is to enjoy an advantage of modesty (cf. *ibid.*, p. 48). Martin assures us it is:

“Well a proponent of the immodest view cannot fault a modest account for failing to capture in its conception of what a sense experience is all those situations that the immodest account deems to be perceptual experiences of a street scene. After all, by immodest lights the kind of experience one has when seeing such a street scene is of just the same kind as any non-perceptual event which is not a perception but still an experience as of a street scene, namely an event with

---

<sup>36</sup> This question applies as much to the wakeful, non-deluded subject as to the dreamer, for whom it is also possible, because actual, that she could be hallucinating.

the properties  $E_1, \dots, E_n$ . Since nothing can be discriminated from itself, the immodest approach will hold that the modest one should agree that these events are indiscriminable from a veridical perception of a street scene and hence are perceptual experiences as of a street scene” (*ibid.*).

Moreover, the minimal conception is not too liberal. It could only be too liberal if cases indiscriminable from perceptions *shouldn't* be thought of as experiences. But that sounds bad. Accordingly, the situation is really that the non-minimal conception needs its extension to coincide with that of the minimal conception. And this leads to further immodesty.

The only way a proponent of the non-minimal conception might secure sameness of extension is by declaring impossible the case in which a situation lacks experience-making features  $E_1, \dots, E_n$  and yet is indiscriminable from the Good cases. With indiscriminability characterized in the manner of Williamson (1991) (more on which presently), this is to declare impossible any case which fails to exhibit  $E_1, \dots, E_n$  and cannot be told apart from the Good cases. Given the proponent of the non-minimal conception needs this, they therefore need that subjects are always in a position to discriminate situations which lack  $E_1, \dots, E_n$  from those which have  $E_1, \dots, E_n$  (50).

And *that* is where the immodesty is unpalatably compounded. To commit to the above is to accept that a subject engaged in careful reflection must always be able not only to detect the experience-making features of an event when they are present but be able to detect their absence as well (Martin, 2004, p. 51). This amounts to a form of infallibilism about experience (*ibid.*). Some philosophers may not mind this consequence of this conception of experience. But if we can accept that we are not infallible with respect to one domain of experience-regarding judgments,<sup>37</sup> then it is a stretch to suppose us infallible with respect to any.

---

<sup>37</sup> Martin does not provide examples of such fallibility in his paper, but Dennett (2005) provides a wealth of them. Consider change blindness (ch. 4.2), the phenomenon wherein substantial changes in our visual fields go unnoticed

The considerations just presented motivate conceiving of perceptual experience *in general* as episodes indiscriminable from the Good case. But it does not yet clearly specify what the Bad case is. We know that whatever else it might be, it is indiscriminable from the Good case. But that is true of the Good case, too, and yet we know more besides about the Good case; we know, namely, that it is explicable as per NR. We may be wondering, then, whether more might similarly be said about the Bad case.

RD's take specifically on the Bad case is this:

**RD<sub>Bad</sub>** The Bad case as of  $\Phi$  fully consists in indiscriminability from the Good cases of  $\Phi$ .

with indiscriminability understood, following most reflective disjunctivists (Martin, 2004, 2006; Fish, 2008, 2009; Brewer, 2011; Soteriou, 2013) along the lines of Williamson (*op. cit.*), on which for  $x$  to be indiscriminable from the  $F$ s is for it to be impossible for an impersonalized subject, reflecting solely on epistemic context, to activate knowledge that  $x$  is not one of the  $F$ s (cf. Sturgeon, 2008, p. 127).

RD<sub>Bad</sub> is first postulated in in Martin (2004). Its first mention comes on the tail of two commitments: first, to the presence of a psychological kind common to cases of perceptions and causally matching hallucinations. This commitment is the result of a concession the disjunctivist, of any stripe, is advised to make in response to an amended version of the argument from hallucination against naïve realism (*ibid.*, p. 54ff). The premise of the argument that the disjunctivist is advised to accept is that where neural conditions are exactly matching—as they

---

by us; or “filling in”, the phenomenon whereby we testify to being able to attend to much more in our visual fields than is really there (65-69); or our penchant for overestimating how far color discrimination extends into the peripheries of our visual fields (41).

are across perceptions and some hallucinations—there is a shared psychological kind (53-54).

Now, we are already conceiving of perceptual experiences generally as exactly those episodes that are indiscriminable from perceptions. The property *being indiscriminable from a perception*, then, is an excellent candidate for the common psychological kind.

The second commitment is a desire to avoid the noted common kind's *screening off* the explanatory role played by perceptions (*ibid.*). That is, if the disjunctivist posits the property *being indiscriminable from a perception* as the psychological kind shared between perceptions and causally matching hallucinations, then there is a danger that it is that property which explains the things we invoke perceptions to explain. Which things are these? Things like perceptions' cognitive effects, and its behavioral consequences, among other things. So, in order that that property not screen off perceptions from explaining all those things we invoke perceptions to explain, it must be that causally matching hallucinations fundamentally consist in the property, whereas perceptions do not. Hence, whereas perceptions fundamentally consist in subjects' standing in irreducible perceptual relations to mind-independent objects/ properties (i.e., are grounded as per NR), causally matching hallucinations fundamentally consist in their indiscriminability from perceptions.<sup>38</sup>

---

<sup>38</sup>Two caveats are in order. First, disjunctivists often note, in the same breath in which they embrace commitment to causally matching hallucinations' fundamentally consisting in indiscriminability from perceptions, that they can remain agnostic as to what non-causally matching hallucinations consist in. This commitment naturally raises a question that I think would needlessly sidetrack the discussion, namely, whether experiences as of supersaturated red are causally matching hallucinations. If not, we might think, then the reflective disjunctivist does not need to say anything about it. We will find out later that the experience is not a causally matching hallucination. But two points count against fussing over this. First, RD says that to be a perceptual experience is to be indiscriminable from a perception. Later we will see that the same considerations which motivate taking the example not to be a problem for RD<sub>Bad</sub> are exactly the considerations which make the experience a problem for RD. So, the reflective disjunctivist should care that any experience fails to gel with the extension it assigns experience. *Especially* given the

RD so characterized, along with its austere take on the Bad case, and NR are the theses with which I will be concerned throughout. Before I recap the two theses' motivations, a word on why I restrict attention to RD from among the other run-of-the-mill disjunctivist theories. First, non-reflective disjunctivist theses are much less popular now than they were before the turn of the millennium. So a discussion of their prospects for accommodating the problem experience would be of limited appeal. Second, RD scores massive points for being far simpler, at least in terms of parsimony, than its non-reflective cousins. If our sympathies lie with disjunctivism at all, we should care that the simplest of the run-of-the-mill theses faces counterexamples. Finally, RD is a cornerstone commitment of a handful of major naïve realist accounts recently defended (Fish, 2009; Brewer, 2011; Soteriou, 2013). Counterexamples against the view ramify for these major works.

To recap, it is semantic, epistemological, or phenomenological considerations which motivate NR. And each of these motivations moreover is based on how worldly character presents itself to subjects, how it is known by subjects, or how it thought about reflection on subjects' contact with mind-independent worldly character. And considerations of epistemological modesty and screening-off, underlain by a presumption of the extensional adequacy of the view's conception of experience, recommend RD.<sup>39</sup>

---

presumption of extensional adequacy central to the RD's conception of experience (points we will return to later on). Second, the imparsimoniousness of a view which does not ground *all* Bad case phenomenology in indiscriminability totally cancels out whatever gains were made with the virtue of modesty. As I said in the main text, I focus on RD at the expense of its non-reflective cognates because of its superior parsimony. For these reasons, I think it best simply to leave RD<sub>Bad</sub> as defined in the main text, and to leave it on the table as a naïve realist resource for handling experience as of supersaturated red.

<sup>39</sup> In a similar vein, Sturgeon (2008) motivates RD by appeal to parsimony. Because everyone, he says, agrees that experiences are in general indiscriminable from corresponding Good cases (119, 122), the theory which grounds the

## 2 *Experience as of supersaturated red*

I will now introduce the experience which poses a challenge to RD, experience as of supersaturated red. And after introducing it, I will briefly motivate taking the experience to be as of a metaphysically impossible property.

Supersaturated red is an afterimage-experience subjects are said to be able to undergo when they stare at a maximally saturated red surface after having fixated on a green one for a longish period of time or after having had green light shone into their eyes for a prolonged period.<sup>40</sup> Fixating for a prolonged time on green, or having green light shone into the eyes, *potentiates* the cells implicated in things' appearing red (and green) to subjects. What this means is the cells undergo a storing up of chemical resources. And this storing up of chemical resources results in the subject's seeing the *maximally* saturated red surface that they subsequently fixate on as being more red than it really is—or could be.<sup>41</sup>

---

Bad case in indiscriminability is at a dialectic advantage vis-à-vis its rivals (122). Crucial to this motivation too is the assumption of RD's extensional adequacy. For if experiences were not in general indiscriminable from corresponding Good cases, then the dialectic advantage would be lost.

<sup>40</sup> I think it is important to note a hesitation we should have with the example. Churchland (2005) says that there can be no afterimage-experience as of supersaturated red (188-9). This is because, in short, experience as of maximally saturated red already involves implicated neurons' *full* activation. Potentiating the cells before turning one's fixation to a maximally saturated red surface will not result in their being pushed beyond full activation, so no experience as of a red more saturated than fully saturated red should result. However, experience of maximally saturated orange, e.g., does not involve implicated cells' full activation. And this allows for the possibility of inducing afterimage-experiences as of an orange more saturated than fully saturated orange. For the reader inclined already to accept that subjects cannot undergo supersaturated red experience, read each use of 'supersaturated red' below as denoting supersaturated orange experience instead.

<sup>41</sup> Here is Hurvich (1981) on the phenomenon:

“If the primary excitation in a small foveal field in an otherwise dark surround is produced by, say, 500 nm, it looks green while the stimulus is on. If we turn the stimulus off and look at a small not-too-bright

‘Maximally’ here denotes that the red surface is saturated to the highest degree. There could not be a more saturated red surface. This is so because the surface instantiates the highest degree of the physical quantity which subvenes saturation. (I say “subvene” rather than “reduces” to remain neutral between reductionism and primitivism about color properties. For detailed discussion of each view, see the following essay.) So, when we experience the surface as being more saturated than maximally, we have an experience as of an impossible property, a degree of saturation beyond maximum. What *grade* of impossibility is this? Let me briefly motivate taking the answer to be “metaphysical”.

First, note that NR is committed to the mind-independence of perceptible qualities, qualities like saturation, hue, brightness, etc. (This much was included in our characterization of the view in section 1.) This means saturation must supervene on an external property of objects that is individuated independently of subjects’ responses to the property. Just above, I said that this property is a quantity, a property that can be exhibited in varying degrees. In perception of maximally saturated red surfaces, the saturation underpinned by the highest (maximal) degree of this quantity shapes the contours of subjects’ experience of that redness.<sup>42</sup> That is, what such

---

achromatic surface, we see a red afterimage. If the afterimage is superimposed on a small red field, we perceive a SUPERSATURATED red (“supersaturated” means more saturated than the saturation of a narrow-band spectral stimulus of the same hue)” (187).

In the following essay, I discuss implicated mechanisms in detail

<sup>42</sup> It is best to remain relatively silent with respect to exactly which quantity this is. Color scientists do not concern themselves with what external properties saturation (or hue, or lightness) supervenes on. According to Pautz (forthcoming), most color scientists do not think colors are mind-independent at all (22). So, there are scant proposals regarding the supervenience base of saturation qua property of external objects. And for that, it is difficult to prove something exhibits the highest possible degree of saturation. Proofs must be indexed to specific reduction proposals. But there is one proposal on which the claim of metaphysical impossibility goes through. On Churchland’s (2007) account, saturation supervenes on the *tilt-angles* of metamers’ *canonical approximation ellipses*. And experiences as of supersaturated hues can be induced over surfaces which exhibit the maximal degree of saturation’s



cases are like for subjects is partly constituted by subjects' perceiving that highest degree. Because the quantity is exhibited to its highest degree in those cases, this makes it so that no other degree of that quantity is available to underpin the saturation presented in the afterimage-experience under discussion. Not, anyway, without violating the assumption of supervenience: if a distinct physical property underpins the presented higher level of saturation, then it is not true that saturation supervenes on the original quantity.

So the experience is as of a metaphysically impossible property. And for being as of the impossible, the experience cannot be accommodated by RD. I explain why that is in the following section.

### *3. Experience as of the impossible as counterexample for RD*

In his (2004), Martin comments that

“we may question whether every conceivable hallucination has a corresponding veridical perception for it to match. Consider an hallucination of an Escher-like scene with an impossible staircase, for example; or the non-perception of Mark Johnston's example of supersaturated red” (80).

Immediately after, he goes on to offer a strategy for accommodating these experiences, which we will look at below. But what matters for us here is why exactly these sorts of cases present a difficulty for RD. We can gather from the quotation that Martin himself thinks the account is committed to there being for any hallucination a corresponding Good experience that “matches”. But not much more than this is offered regarding what exactly the problem is. I want

---

supervenience base (if Churchland, 2005, is anything to go on—particularly pp. 188-90). I discuss this proposal in more detail in the final essay. And I elaborate the point just made in the Appendix to this essay.

to fill this gap and offer an interpretation of how exactly experiences like experience as of supersaturated red generate problems for RD.<sup>43</sup>

Because we are considering experience as of the impossible, the experiences will not be perceptions. This makes it initially very intuitive that experience as of the property is Bad. We will work with this assumption for now, but we will drop it later when we look at what sorts of things the naïve realist might say about experience as of the impossible. And to further simplify discussion, I will restrict attention just to RD's take on the nature of the Bad case as of *property-instances*.<sup>44</sup> RD says Bad cases as of *Fness* consists in its being impossible for subjects to know the experience to be distinct from the Good cases of *Fness*.

RD disallows experience as of metaphysically impossible *Fness* for the following two-part reason.<sup>45</sup> To start, experience as of metaphysically impossible *Fness* (like experience as of supersaturated red) could only be *vacuously* indiscriminable from perceptions of *Fness*. Why? It is true that there is no world in which subjects know (possible) experience as of *Fness* to be distinct from metaphysically impossible perceptions of *Fness*. *But this is only because* they are given *no opportunity* throughout modal space to fail to know them to be distinct. Contrast this with the non-vacuous case, where experience as of metaphysically possible, perceptible *Gness* cannot be known to be distinct from perceptions of *Gness*. This is a substantial fact because there exist perceptions of *Gness* in modal space from which subjects genuinely fail to distinguish their

---

<sup>43</sup> Siegel (2004, 2008), who discusses Escher scenes as counterexamples to RD, also leaves unsaid why exactly experiences as of the impossible serve as counterexamples. However, as I explain in note 46, her case, experience of Escher scenes, works importantly differently than mine.

<sup>44</sup> Restricting attention to property-instances in this way should be fine. I know of no naïve realists today who deny that subjects are visually related to property-instances in addition to objects. Brewer (2011) does not explicitly say that subjects can stand in visual relations to property instances. But in a recent updating (2018) of his official position, he does say subjects can do this.

<sup>45</sup> I leave open that there may be other reasons.

experiences. This marks the beginning of the problem experience as of impossible  $F_{ness}$  poses for RD.

The problem is compounded when we note the following. Given the metaphysical impossibility of perceptions of  $F_{ness}$ , in no world does the subject know her experience as of  $F_{ness}$  to be distinct from perceptions of  $F_{ness}$ . If experiences as of metaphysically impossible  $F_{ness}$  are metaphysically possible, then it must be true of at least one experience that it cannot be known distinct from impossible perceptions of  $F_{ness}$ , or else the experience is assigned no phenomenal nature. But here is the issue. It is true of *all* experiences that they cannot be known distinct from metaphysically impossible perceptions of  $F_{ness}$  if it is true of any. After all, there is no basis on which to declare experience as of possible, perceptible  $G_{ness}$  distinct from *perceptions* of metaphysically impossible  $F_{ness}$ : the perceptions exist in no world. The tempting rejoinder we must avoid is that we know the (impossible) perception's phenomenal nature on the basis of knowing the nature of experience as of  $F_{ness}$ . Here is why we cannot say this. Disjunctivists are committed to the Bad case as of  $F_{ness}$  having a distinct phenomenal nature from the Good cases of  $F_{ness}$ . This has largely to do with the notion of indiscriminability at play. A subject's being unable to know  $x$  distinct from the  $F$ s does not license the inference that  $x$  is  $F$ . So, even if we grant subjects cannot tell possible experience as of  $F_{ness}$  apart from impossible perceptions of it, this does not license the inference that the impossible perception therefore shares a phenomenal nature with the experience as of  $F_{ness}$ . This is all just to bring out that by thinking of the impossible perception's nature as ably fixed by reference to the possible experience as of  $F_{ness}$  is to get the order of individuation the wrong way around. It is the perception's nature which individuates the nature of the non-perception.

An analogy should make this last point clear. Suppose we thought it true to say that the regally dressed man at the Palais Bourbon, the man surrounded by reporters and cameras, could not be told apart, strictly on the basis of vision, from the King of France. If we thought it appropriate

to say this, then should not go on to say that what it is to look like an individual is to be unable to be told apart, strictly on the basis of vision, from that individual. This because drastically different-looking individuals could easily simultaneously be unable to be told apart, just on the basis of vision, from the King of France. The casually dressed man strolling through Montmartre might be the King of France trying to avoid media attention; the *other* regally dressed man at the Palais Bourbon cannot be visually told apart from the King of France if the first man cannot; and so on and on. Everyone is just as good a candidate as the next for being unable to be visually told apart from the King of France. And you could never justifiably say that some one person does not look like the King of France because, there being no King of France, there would be nothing on which to base the denial. And, so, if we identify looking like an individual with an inability to be told apart, then this give us the absurd consequence that drastically different-looking individuals look like the same individual. (Let me nip in the bud the suggestion that drastically different-looking individuals *can* look like a single individual, so long as they do so in different respects. These respects are impossible to specify in the King of France case.)

What we see, then, is that if experience as of the metaphysically impossible is possible, then RD is in trouble because the experience is only accommodated by RD vacuously. And, what is worse, RD assigns every experience the nature it assigns experience as of the metaphysically impossible.<sup>46</sup>

---

<sup>46</sup> *N.b.* the above covers just *one* flavor of experience as of the impossible: the case where the impossibility of *Fness* is not known on the basis of reflection alone. I discuss just this case because experience as of supersaturated red fits this mold. But there is also this flavor: an impossibility is presented in experience and reflection on context reveals the impossibility. In such a case, reflection on context rules out that there are any such perceptions and, so, the experience is known to be distinct from any corresponding such perceptions (given their known nonexistence). This is a problem for RD because RD then assigns the experience *no* nature. This happens with experience as of Escher scenes (cf. Siegel, 2004, 2008). Below we will see how Martin handles these experiences.

And because experience as of supersaturated red is experience as of the metaphysically impossible, prima facie RD cannot accommodate the experience. However, this is just prima facie. In the next section, we will look at the extant attempts in the disjunctivist literature to say more to accommodate experience as of supersaturated red.

#### *4 Attempted strategies of accommodating supersaturated red*

We will now look at the two attempts in the literature to accommodate experience as of supersaturated red. We will look at Martin's (2004) first. Martin attempts to handle the experience by arguing that it should be thought of as decomposing into constituent elements each of which is unproblematic from the perspective of RD. We will look second at Kalderon's (2011). Kalderon handles the experience by arguing that it should be thought of as being grounded in the afterimage's objectively looking supersaturated red.

##### *4.1 Martin: treat supersaturated as decomposing into kosher constituent elements*

Above, I shared a quotation from Martin (2004) wherein he presents supersaturated red as constituting a prima facie problem for RD. Immediately after the shared quotation, Martin recommends accommodating the problem experience by treating them like so:

“So how is the account to be extended to these cases? One move would be to discuss not experiences per se but rather the various aspects of an experience, the different entities which one can experience and the ways in which they can appear to one....To generalise the account, we would need to fix on the various aspects of a state of perceptual awareness, the ways in which it may be the same or different from other such states of awareness....The beginning of an approach to partial hallucinations is then to explain those aspects of the experience which are not perceptual in terms of that aspect of experience's indiscriminability from the corresponding aspect of a perceptual awareness of that element. In turn, one may seek to explain certain impossible

experiences not by direct appeal to the idea of a veridical perception of that scene, but rather by explaining how an experience with each of the constituent elements is indiscriminable in that respect from a perception of that element” (2004, pp. 80-81).

The injunction is to break perceptual experiences down into constituent elements. Some perceptual experiences, taken wholesale, might be as of scenes with no corresponding Good case. But the hope is that when we break down these experiences into their constituent elements, each element of the experience will be the sort of thing that can take part in a perceptual experience which is, we might say, *fully* Good. That is, each element of the perception as of an impossible scene can feature in an experience no element of which is inappropriate given the scene presented. (My phrasing of Martin’s recommendation may sound more complicated than Martin’s own phrasing. But the complication is called for. Martin’s proposal, taken at face value, would see us analyzing aspects of hallucinatory experiences in terms of Good experience of those very aspects. But there is no such thing as Good experience of aspects of hallucinations.)

How, then, might this strategy apply to cases? Martin never specifies. But we can make the strategy plausible by conjecturing intuitive applications. Consider first the Escher case. A natural decomposition sees the constituent elements of the experience as respectively underpinning the presentation of individual flights of stairs, of the connections of pairs of flights, of whether individual flights are ascending or descending, and so on. For the sake of argument, we can grant that this strategy is not inappropriate for cases like visual experience as of impossible scenes, like the Escher staircase.<sup>47</sup> What we will soon see is that even if the strategy accommodates cases like those, it does not account for supersaturated red.

For an example that gets us closer to our target, consider afterimage-experience as of navy blue. If we wanted, it seems we should be able to say that the experience decomposes into constituent

---

<sup>47</sup> See Siegel (2004, fn. 6) for the complaint that it is ad hoc and tailored to one example.

hue and hue-brightness elements: into blue (hue) and dark (brightness). Why? Because we can say of the experience that it is the *same* as other experiences as of blue, those as of sky blue and cerulean, for instance, in respect of the hue presented. But it is *different* from experience as of sky blue and cerulean in respect of brightness (and maybe also saturation). Hue and brightness are dimensions along which the experience might be similar or dissimilar to other experiences. So they seem to fit the bill. It is important to note, however, that if color experience does *not* intuitively break down into constituent elements like these, then Martin's proposed treatment of the problem experience seems hopeless.

Now, although the above manner of speaking of color experiences which differ along dimensions of hue and brightness (and saturation) is cogent, there is no natural strategy for breaking down supersaturated red into constituent elements that gets rid of the problematic element. That problematic element, recall, is supersaturation. Supersaturated red experience is like experience as of fully-saturated crimson and it is like experience as of low-saturated red: it is like both experiences in respect of hue; it may be like the latter, but it is not like the former in terms of brightness; and it is like neither in respect of saturation (even though it is *more* like the former than the latter in terms of saturation). So, it shares in common with color-experience in general that there are elements of hue, brightness, and saturation into which it might decompose. But it is one element that is problematic in the experience: to wit, saturation. And adversion to none of the above elements dissolves that problematic element. Saturation does not plausibly decompose. (In color science textbooks, it is one of the fundamental dimensions along which color experiences differ—the others being hue and brightness. See, for instance, Kuehni, 2003.)

Accordingly, the current strategy for handling supersaturated red does not succeed. I see no obvious bar to treating color experiences as decomposing into distinct elements of hue, saturation, and brightness. But that aspect of experience as of supersaturated red which involves the presentation as of beyond-maximal saturation (of, i.e., supersaturation) does not decompose.

When experience involves the presentation as of beyond-maximal saturation, we are left without a plausible means of further decomposing the experience which would root out the problematic element.

#### *4.2 Kalderon: treat the afterimage as instantiating a supersaturated red look*

We have just seen that RD cannot handle supersaturated red by appealing to that resource which belongs uniquely to it, namely, indiscriminability. In order for indiscriminability facts to ground what experience of supersaturated red is like, the impossible saturation of the experience would need to have plausibly decomposed. But it did not. This failure to accommodate experience as of supersaturated red with those resources unique to RD, however, immediately ramify for RD. It would be a problem that indiscriminability facts cannot ground experience as of supersaturated red if we were forced to think of the experience as Bad. This because all RD has to say about Bad experience is that it is indiscriminable from Goodness. But the possibility is left open that supersaturated red is Good. If supersaturated red is Good, then, accordingly with RD, it gets handled with the resources of NR. The following strategy we will look at does this. It takes experience as of supersaturated red to be Good. If it is viable, then supersaturated red is not a problem for RD. I will argue, however, that the strategy does not succeed.

Kalderon (2011) argues for a naïve realist treatment of supersaturated red. In order for supersaturated red to fall within NR's purview, we just noted, it must be Good. This may sound like a confusing idea: afterimages intuitively are not Good—their phenomenal natures intuitively do not consist in subjects' being related to mind-independent properties. In order to understand how experience as of supersaturated red might be Good, we need to introduce the notion of *objective looks*.<sup>48</sup> In the following quotation, Kalderon gives us a sense of what looks are:

---

<sup>48</sup> See also Brewer (2011); Martin (2010); Travis (2004) for naïve realist uses of looks.



“If experience is relational, then there must be something to which the subject is related. What, then, are we related to in cases where there is a contrast between appearance and reality?

Something can look *F* without being *F*. Here we have a contrast with appearance and reality. To look *F*, though, is to have a certain look. A look is a way for things to be. It is a feature of things that grounds objective similarities (in this case between the thing seen and certain things which are genuinely *F*)” (773).

And, so, in normal viewing circumstances, *F* things look *F*: red things look red; square things look square; moving (middle-sized) objects look to be moving. But these are not the cases Kalderon is interested in. Kalderon, as we see in the quotation, is concerned with things whose looks are in conflict with how they otherwise are. The more interesting examples of things which objectively look certain ways, then, are: the straight stick half-submerged in water which, due to the light it reflects being refracted by the water, looks bent (767); the white bead that in blue lighting looks blue (761-62). In these cases, when standpoints are suitably specified, these objects possess objective looks. The ways objects look we indirectly specify by way of making looks-statements which are comparative (772). That is, we specify the way-a-thing-looks by comparing it to things which, in suitably specified circumstances, genuinely are that way. The submerged stick looks like a bent stick (out of water); the white bead looks like a blue bead (under normal illumination)—and all the *F* objects which look *F* look like *F* things (in suitably specified circumstances).

The broader aim of Kalderon’s (2011) is to argue that *all* putative visual illusions can be handled this same way. All putative visual illusions are, if surprisingly, visual Good cases. Recall that visual Good cases involve subjects’ being irreducibly visually related to mind-independent entities/features of their environments. And here, the relevant mind-independent features are looks (773). In what philosophers traditionally call visual illusions, what is really going on is that subjects are visually related to look-features objectively possessed by the seen objects. Because the objects/properties in these (and all) cases look to subjects in accord with what look-features

the objects/properties possess, Kalderon concludes that there are no illusions properly so-called.<sup>49</sup>

In defending this idea, Kalderon anticipates an objection to the effect that supersaturated red could not be a case of a subject's being visually related to an objective supersaturated red look because nothing could be supersaturated red—'could' here (p. 771) deserving special emphasis. This brings us to our second proposal. With the notion of objective looks on the table, Kalderon proposes the following accommodation of experience as of supersaturated red: the experience should be handled by thinking of the afterimage as being in possession of a supersaturated red look (*ibid.*). That is, experience as of supersaturated red consists in subjects' being visually related to afterimages which objectively instantiate the supersaturated red look.

There are two problems I want to raise for this treatment of supersaturated red. The first has to do with appealing to afterimages as being the bearers of looks. The second has to do with actual things' looking like impossible ones. Let us look at these problems in turn.

Here is the problem with recommending we think of afterimages as being the bearers of the relevant look-feature. It only makes sense for the naïve realist to appeal to looks in their treatments of what grounds Good phenomenal character if looks are possessed by the objects/properties of the mind-independent world. Special emphasis is due "mind-independent". If it is not mind-independent entities which have looks, then we have given up on NR. But afterimages are in the first instance experiences. And for that reason they cannot be the bearers of objective looks.

---

<sup>49</sup> In one particularly interesting case, a circle that, when stationary, is black and white all over looks to subjects, when it is spun, variously-hued. Kalderon says this multicolored-look too is a feature of the spinning circle and is no illusion. The spinning circle itself possesses the look feature that subjects come to be visually related to.

This issue could be overcome if there were something in the mind-independent scene which might plausibly have the supersaturated red look. Perhaps a plausible suggestion could be made in the vein by appealing to that resource of NR emphasized in section 1, namely, standpoints. Perhaps by making explicit what exactly are the viewing circumstances implicated in experience as of supersaturated red, the naïve realist will more easily be able to specify which features of the mind-independent scene possess the relevant look. For instance, we might say that in the very specific circumstances wherein subjects' Green/Red neurons are potentiated and subjects subsequently fixate on a red surface, it is not the afterimage but the (maximally saturated) red surface which really looks supersaturated red.<sup>50</sup> We are no longer grounding afterimage-experience as of supersaturated red in the afterimage's having a look. It is a part of the mind-independent environment, indexed to neuronally specified viewing circumstances, which now has the look.

Here is a concern for this version of Kalderon's proposal. One might worry that it too lopsidedly grounds the experience on the *subject-side* of the naïve realist relation. Part of what makes NR so appealing as a thesis of the phenomenal character of the Good case is that it can intuitively account for so much of experience without having to invoke intrinsic features of subjects. This much was made apparent in our discussion of each of its motivations. The view is motivated just by appeal to the way worldly character presents itself, is known, or is thought about. But qualitative features of the subject is virtually erased in discussions of why we should endorse NR.

We can do more than appeal just to NR's motivations here. Consider this following scenario, which, by the lights of the account under consideration, gets classified as Good:

A vision scientist has very cleverly designed a short video on which a subject is to fixate.

The video is designed to do two things: make it so that a multitude of afterimages are

---

<sup>50</sup> This sort of response was suggested to me by Craig French (in conversation).

experienced after fixation; and achieve the effect characteristic of the waterfall illusion, whereby stationary objects appear moving. After fixation, the subject looks at a white wall, and it seems to her that she is watching a scene from Disney's *Fantasia*.

It is too much of a stretch to call this episode Good. What is more plausible is that the scientist has caused the subject to hallucinate. But if neural goings on are not inappropriately appealed to in our specification of the standpoint to which Good cases are indexed, then we have no obvious grounds for denying the movie-experience is Good. The present worry does not give us a knockdown objection to this version of Kalderon's proposal. But it should lower our confidence in it.

There is a worry that persists even if we grant the above amendment. There remains a problem with the noted impossibility of supersaturated red (Kalderon, 2011, p. 771). One consequence of the impossibility of supersaturated red is that no supersaturated surfaces exist actually. And, so, Kalderon says on behalf of his account of supersaturated red that

“[according to the relevant sense of comparison at work,] the afterimage in the present circumstances looks the way a supersaturated red thing would look like in some other, contextually specified, circumstance. However, supersaturated red is a missing shade. Nothing is supersaturated red. So there is nothing that we are comparing the after image to. If we have failed to make a comparison then we have failed to indirectly specify the way of looking by means of that comparison. However, we can qualify something be means of a comparison without the standard of comparison existing” (772).

This comment picks up on the failure of anything actual to be supersaturated, this failure being a consequence of the noted impossibility. And we see that Kalderon does not think that the property's being a “missing shade” is any bar to its being a standard of comparison. But the comment downplays the *impossibility* of an object's instantiating supersaturated red. (Recall that we are taking supersaturated hues to be metaphysically impossible, which decision is defended in

the Appendix.) To see this, note that at the start of the quotation, Kalderon says the afterimage looks the way a supersaturated red thing *would look* in suitably specified circumstances. But intuitively, propositions to the effect that “some  $x$  would  $\varphi$  in circumstances  $C$ ” are made true by what goes on in those worlds in which  $x$   $\varphi$ 's in  $C$ . But, given the impossibility of supersaturated red, there are no worlds in which supersaturated red (our value for the schematic letter ‘ $x$ ’) does anything at all in any circumstances at all. So, the statement “the afterimage would look supersaturated red...” is false or truth-valueless on a semantics which does not appeal to impossible worlds as truth-makers of modal claims. And this matters because if we are not antecedently attracted to a view on which impossible worlds are the truth-makers of modal claims like the above, then to go for such a view just on being faced with the present challenge would be ad hoc.

But we can bracket the issue of whether the disjunctivist is antecedently attracted to impossible worlds. There is a further set of questions about which the disjunctivist is at risk of making *ad hoc* claims, questions having to do with how impossible worlds are to be related to the actual world in terms of closeness. On the usual (Stalnaker-Lewis) semantics, counterfactuals are true only if in all the closest possible worlds in which the counterfactual's antecedent obtains, the consequent does too. And so, if it is said that “ $x$  would have  $\varphi$ 'ed had circumstances  $C$  obtained,” to find out whether this is true, we look to the closest worlds in which  $C$  obtains and check whether  $x$   $\varphi$ 's there. There is lots of debate about what determines closeness relations among possible worlds. It a less explored issue, though there has been much said nonetheless, what determines relations of closeness among impossible worlds. But how they are to be ordered is of definitive importance to the disjunctivist who wants to say that impossible worlds are what make impossible counterfactuals true. This is because the statement “where  $C$  obtains, supersaturated red looks like *this*” (where ‘ $C$ ’ denotes the circumstances we have so far worked with in which G/R neurons are potentiated before the subject looks at a maximally saturated red surface, and where ‘this’ demonstrates the afterimage-experience that the subject is currently

undergoing) is true only if in all of the *closest impossible* worlds in which C obtains, supersaturated red does look that way. Here is the problem. There are also impossible worlds at which supersaturated red, in those same circumstances, looks white, or looks fully saturated red, or looks many other ways.<sup>51</sup> And if supersaturated red would look white in the relevant circumstances, then we could also say that experience as of supersaturated red involves being visually related to something with a white-look. But it is experience as of white that is grounded in visual relations to white-looks. And experience as of supersaturated red is decidedly not like experience as of white. The disjunctivist needs to say something, then, and something non-*ad hoc*, about why the worlds she needs to be closest are indeed closer than the worlds just mentioned.

What I would anticipate she could say here is that because it is more intuitive that supersaturated red would look like *this* rather than, say, like fully saturated red, we should say that the impossible world in which it looks like *this* is closer than the impossible worlds in which it looks like fully saturated red. But nothing has been said to make it plausible that intuitions are a good guide to which (impossible) worlds really are closer. Moreover, we are given no reason to suspect that the disjunctivist's intuitions here are not liable simply to track what she needs to be the case rather than what is the case. To elaborate, suppose some bit of metaphysical theory depended on its being true that elements with atomic number less than that of hydrogen should behave thus and so. Such elements are arguably impossible. I do not think our intuitions should be trusted in this case if reflection on the issue yields that it is more intuitive that they behave that way rather than some other way if our theory depends on their behaving in that first way. Intuitions may not be our best guide here.

---

<sup>51</sup> Because looks are mind-independent properties, they are separable in this way from experience's being as of the look.

We have seen a few reasons to be doubtful that Kalderon's handling of supersaturated red succeeds. First, his account had us attributing objective looks to what should count as mind-dependent entities (afterimages should count as mind-dependent, anyway, until they have been successfully accounted for with the resources of NR). This misplacement of objective looks could be avoided if we sufficiently specified the standpoints from which mind-independent items have the relevant looks. But when standpoints are sufficiently specified, we are stuck with labelling Good visual episodes that seem far from it. This was the first set of issues. We arrived at the next set of issues by noting certain surprising semantic commitments the disjunctivist would have to make in the case that she wants to say how supersaturated red would look (if it existed). She needs to accept that impossible worlds make certain counterfactual statements true. And if she was not antecedently attracted to all that, then the new commitment is ad hoc. She moreover needs to commit to certain ordering relations among those impossible worlds, to avoid making it the case that supersaturated red looks white, or looks normally saturated red, or what have you.

It is crucial I note, before we proceed, that Kalderon's failing to convincingly accommodate supersaturated red does not ramify for his recommended handling of illusions more generally. Kalderon, for all I will have said, may still have successfully shown that all of the standard cases of what we call illusions are really cases of subjects' being visually related to objective looks possessed by features of the mind-independent scene. I do not think the failure of afterimage-experiences as of supersaturated red to be grounded in objective looks undermines his cause. This has to do with the fact that Kalderon already needs to say something different about hallucination. He does not say hallucinations are grounded in looks. He could, accordingly, categorize afterimage-experience as of supersaturated hues as hallucinations. And there is already precedent in the literature for categorizing afterimage-experiences this way (e.g., Brewer, 2011, p. 115). If afterimage-experiences as of supersaturated hues are appropriately categorized as

hallucinations, then there will be much less temptation to think of them as visual illusions. (Or rather, he should say that afterimage-experiences in general are a kind of hallucination.)

What I have argued so far is that the extant attempts by disjunctivists/naïve realists to accommodate experience as of supersaturated red do not succeed. The first case we looked at, Martin's, we saw was unable to accommodate the experience because there was no plausible story to tell whereunder the experience decomposes into constituent elements each of which is kosher from the perspective of RD. The second attempt, Kalderon's, was unable to accommodate the experience because, first, it needed to load too much of the experience's grounds on the subject-side of the naïve realist relation; and second, because it required impossible worlds as the truth-makers of the looks-statements it needs to appeal to in its account of what grounds the experience, which entrained its own set of forced hands.

In the next section, we will address whether the disjunctivist could simply deny the problem. What if the disjunctivist denies the phenomenology we have hitherto assumed to be correct: what if she denies that the experience is genuinely experience of beyond-maximal saturation?

#### *5 An unpromising way around the problem*

In presenting the challenge experience as of supersaturated red poses for RD, we have so far relied on the assumption that the experience is genuinely as of beyond-maximal saturation. It proved a problem for both Martin and Kalderon for this reason. Before concluding, I want to explain why this assumption is a good one. The disjunctivist should not simply deny that supersaturated red poses a genuine problem for her.

#### *Deny the experience is genuinely as of beyond-maximal saturation*

What if the disjunctivist responded to experience as of supersaturated red by denying that the experience is genuinely as of beyond-maximal saturation? What if the disjunctivist went on to give a reply like this:



Experience as of supersaturated red involves the presentation as of a patch of red more saturated than a very highly saturated red surround. But it does not involve the presentation as of a red more saturated than a maximally saturated red surround. What motivates this denial? Well, it does not seem in the first place that on being presented a maximally saturated red surface, subjects experience the surface *as* instantiating maximal saturation. If it did, then subjects should be good at answering, when queried, whether a patch of color is maximally saturated or not. We could empirically test for this competence, or lack thereof, by asking subjects whether red surfaces shown to them are maximally saturated. We could track subjects' accuracy at this performance by showing them red surfaces just shy of full saturation, and others that are fully saturated, etc.<sup>52</sup> If subjects show poor accuracy with respect to the task, then we should conclude that subjects are poor estimators of maximal saturation. And from this conclusion we could make an inference to the best explanation that this is because their experience involves the presentation of only very determinable degrees of saturation. These would include attributions like "very highly saturated", or "highly saturated", and suchlike determinables.

And in the case that that is right, we can say that there is nothing impossible involved in experience as of so-called supersaturated red. What the experience really involves is the presentation as of a red surround exhibiting the determinable "very high saturation" with

---

<sup>52</sup> This is not a test I know color scientists to conduct. (If not just for the reason that it seems to assume the mind-independence of saturation. As I noted above, most color scientists do not believe colors and color-properties to be mind-independent.) What is common is that color scientists will investigate how fine are the discriminations subjects can make between steps in saturation. But subjects are never asked to judge whether something is maximally saturated. The judgments they are asked to make, rather, involve comparing levels of saturation.

a patch inside that is presented as being of higher saturation. And Good cases of this do indeed exist. Hence, the problem is resolved.

The reason I balk at such a proposal is that it sounds like it probably gives the wrong results for experts. Consider the psychophysicist who deals extensively with maximally saturated surfaces. If she asks for a patch of maximal saturation and is passed one with very high, but less than maximal, saturation, she would know it was not what she had asked for, and she would know this strictly by sight (and without, moreover, needing to compare it to other patches). We could again make an inference to the best explanation, this time regarding her experience's involving the presentation as of determinate values of maximal saturation (or that the determinables involved are significantly more fine-grained than the above ones). Her experience would be as of beyond-maximal saturation and would be the experience which proved problematic for RD. Such an expert should not strike us as a stretch. She is not so unlike the trained musician who has perfect pitch. He can tell, strictly on the basis of audition, which key he hears, and he can do so without having to hear other keys before or after. Such a person easily conceivably could detect whatever the highest human-discriminable pitch is, without having to hear others before or after to base a comparison. Our psychophysicist is like this.

And the disjunctivist is not advised to insist, for instance, that there must be *some* level of grain (some saturation-step size) at which even the expert cannot distinguish maximal saturation from whatever level of saturation, at that same level of grain, is just shy of maximum. Though it is probably true that there is such a level of grain, this does not help the disjunctivist if the difference between the afterimage's apparent saturation outstrips whatever constitutes a single step in that level of grain. To illustrate, suppose the expert cannot tell, just by looking, whether a color patch is maximally saturated or just 99% of full saturation. And for that reason we say her experience presents any maximally saturated patch (or any patch the saturation of which is just shy of maximum) as inhabiting some region between full and 99% saturation. This does not

matter if the afterimage appears to be, say, 2% more saturated than the maximally saturated patch. Because the afterimage in that case would still appear beyond maximally saturated.

Accordingly, we are not advised to hold out hope for the experience's failing to be as of what is genuinely metaphysically impossible.<sup>53</sup>

## 6 Moral

In a compelling new paper, Bill Fish (2021) argues it is helpful to view competing philosophic theories of perceptual experience as bearing deep affinities to Lakatosian (scientific) research programmes, these consisting of, first, a *hard core* of essential commitments which cannot be abandoned without thereby abandoning the programme, and, second, a *protective belt* of potentially refutable claims which connect the core to the world (*ibid.*, p. 33). One of the lessons to draw from these deep affinities, Fish argues, is that devotees of any one research programme are entitled not to abandon the programme in the face of counterexamples. Instead, counterexamples should be treated as puzzle cases to be dealt with as the programme matures, after which time the puzzle cases should be able to be handled. In the scientific case, we saw this happen, for instance, with Darwinism: the received geological wisdom contemporaneous with Darwin's postulation of evolution-by-natural-selection was that the planet was not old enough for Darwin to be right. Instead of abandoning Darwinism, however, which would have been over-hasty, the issue was set to the side. Good thing, because as we know now, the geological time frames match up with Darwinism fine.

---

<sup>53</sup> It warrants emphasis that supersaturated red is metaphysically impossible *given the assumption* that color properties, like saturation, are response-independent (among further auxiliary assumptions). It is no aim of this paper to insist that experience as of supersaturated red unqualifiedly puts human subjects in touch with the metaphysically impossible.

I am sympathetic with much of what Fish argues in that paper. As it applies to the present context, I am sympathetic with that idea that if experience as of supersaturated red were a mere counterexample to RD, it would not be inappropriate for the proponent of RD to set the case aside in the hopes that future developments of theory would furnish it with the resources to inoculate the example.

But here's the rub. There are no *mere* counterexamples to RD. Recalling section 1, counterexamples to RD cut against the theory with a double-edge. I have so far made explicit the cut in one direction: experiences as of the metaphysically impossible (for instance) shows that RD is not extensionally adequate. But *for doing that*, it cuts against the theory in this further way, intimated in section 1: it undermines RD's very motivation. That motivation was appeal to epistemological modesty (and on modesty's back, screening-off), where a guiding presumption of the discussion of epistemological modesty was that conceiving of experience in terms of indiscriminability runs no risk of getting experience's extension wrong (see again Martin, 2004, p. 50).

What Fish does not discuss in his (*op. cit.*) is how philosophic research programmes can be undermined by way of having their motivations undercut. And by way of motivation, RD, we saw, rests on a thin reed. Accordingly, it *would* be inappropriate for the proponent of RD to appeal to the ideas of Fish (*op. cit.*) and deny that RD needs (today) to be extensionally adequate.

But even supposing the disjunctivist can get over the awkwardness of flouting extensional adequacy, there is this bigger, independent problem with flouting it. If the proponent of RD allows that some experiences are not indiscriminable from perceptions (and that in this way the extension the theory assigns experience comes up inadequate), then she will have to account for subjects' abilities to identify supersaturated red experience as experience. Because it will not be true that experience as of supersaturated red is indiscriminable from corresponding perceptions, she will have to posit some mechanism/capacity that explains how subjects know the experience

to be an experience. And this will be a “positive” piece of knowledge (Martin, 2004, p. 47). And like that, we lose the advantage of modesty. It will no longer be modest to avoid appeal to positive states of knowledge if we have anyway to appeal to them.

What experience as of supersaturated red does, then, is strip us of a reason to accept RD in the first place.<sup>54</sup> This ramifies for how the reflective disjunctivist should respond to the problem posed by the experience. Because the experience undermines RD’s very motivation, it is not clear whether the proponent of RD is advised to search (especially hard) for strategies or maneuvers that might save RD. If RD had been motivated by appeal to some other explanatory virtues, things could be otherwise. But as it stands, experience as of the impossible knocks out the only leg on which RD stands.

But I want to hedge things a bit and end on a more concessive note on behalf of RD. What the foregoing discussion does show is that not all experiences (or experience elements) are indiscriminable from perceptions (or perception elements). And for showing that, it also shows that an epistemologically modest take on the phenomenal nature of the Bad case will not be cashed out in terms of indiscriminability. But what I have not shown is that there is *no* extensionally adequate account of that nature in the offing which explicates it in terms of *what subjects cannot know*. If there is, then that account would allow for our providing an

---

<sup>54</sup> Note that this is true of any counterexample to RD, and there may be more. See, e.g., Smith (2008) for a battery of further proposed counterexamples. I focus, however, just on supersaturated red because it is a challenge for RD that proponents of the view themselves have acknowledged.

Moreover, the problem presented by the experience is compounded by the following observations. i) The experience-type we have focused on is a member of a class of problem experiences (as of supersaturation). And ii) there are other exotic color experiences the disjunctivist may have similar difficulties with, like red-green (Macpherson, *op. cit.*), or the stygian hues (Churchland, 2005; cf. also the essay following this one). But focusing on how, or whether, these further experience-types cause troubles for RD would have detracted from the focus of the present discussion.

epistemologically modest account of the Bad case. If the proponent of RD found one such “negative epistemic” fit for the role of explicating the phenomenal nature of the Bad case, then she will have preserved RD, if not in letter, at least in spirit. So perhaps that is exactly the sort of phenomena for which the proponent of RD should be on the lookout.

## 7 Conclusion

Experience as of supersaturated red presents a challenge, an acknowledged one, for RD that its proponents have yet to successfully discharge. RD, as currently formulated, does not have the resources to handle elements of experience that could never contribute to a fully Good experience.<sup>55</sup> And for falling outside the extension RD assigns experience, experience as of supersaturated red simultaneously kicks out the one leg on which RD was propped: epistemological modesty. This may prescribe abandoning RD. However, what I have not shown is that there is no extensionally adequate account of perceptual experience in the offing which characterizes experience in terms of *what subjects cannot know*. If there is, then that account would allow for our providing an epistemologically modest account of experience. If the proponent of RD found one such “negative epistemic” fit for the role of grounding experience, then she will have preserved RD, if not in letter, at least in spirit. Perhaps, then, that is exactly the sort of phenomena for which the proponent of RD should be on the lookout.

---

<sup>55</sup> A naïve realist response I did not consider is that of denying responsibility for having to account for experience as of supersaturated red. Masrour (2020) says afterimage experiences *in general* are “pseudo-perceptual” (751) and that, therefore, the perceptual theorist does not need to account for them. But that is implausible for being too pessimistic about the possibility of making computer-generated images subjects confuse for their own afterimages. Masrour must think that experiences of such computer-generated *tromp l’oeil* are not perceptions. If they are perceptions, then afterimage-experiences are indiscriminable from them. (Cf. Lycan, 2019, who argues for the possibility of making computer-generated images that resemble phosphene-experience.) Alternatively, we may just think afterimages are like perceptions of transparent films.



## Appendix

Color scientists do not concern themselves with what saturation (or hue, or lightness) *reduces to*. Most color scientists, as it happens, do not think colors are mind-independent at all (Pautz, 2020, p. 22). Accordingly, it is difficult to prove when something exhibits the highest possible degree of saturation. In order to prove that something exhibits the highest possible degree of saturation, we need to know what that mind-independent quantity is to which saturation reduces. In this appendix, we will see how one argument for the impossibility of the supersaturated hues goes. The argument depends on the proposed reduction-base of saturation to be found in Churchland (2006). *N.b* that Churchland's proposal is not without blemish: Kuehni and Hardin (2010) argue that it attributes levels of saturation which conflict with normal subjects' reports (87).

Churchland's hypothesis is that the properties represented in color experience are what he calls the *canonical approximations* of metamers (2007, sect. III). (Metamers being, roughly, differences in a color's reflectance that subjects cannot detect. Metamers are rife and each color's associated metamers comprise motley sets.) To find a metamer's (reflectance's) canonical approximation (CA), we must first take a graphic representation of a particular reflectance. As may be familiar, these are rectangular graphs the x-axes of which denote electromagnetic radiation wavelength, starting with 400nm and ending at 700nm, and the y-axes of which denote how strongly radiation of a certain wavelength is reflected. One is to roll these rectangular graphs into a cylinder so that the beginning of the x-axis (400nm) meets its end (700nm). (If you have a paper graphic-representation of a reflectance, you can roll it into such a cylinder by cutting the graph out of the page.) The CA of the reflectance is a planar slice, an ellipse, through the resultant cylinder which meets the following conditions:

- “1. the altitude of the ellipse must be such that the total area  $A$  above the [slice], but below the several upper reaches of the target reflectance profile, is *equal* to the total area  $B$  beneath the [slice], but above the several lower reaches of the target reflectance profile.



(This condition guarantees that the *total area* under the target reflectance profile equals the total area under the [slice].)

2. The angle by which the [slice] is tilted away from the horizontal plane, and the rotational or compass-heading positions of its upper extreme, must be such as to *minimize* the magnitude of the two areas *A* and *B*. (This condition guarantees that the [slice] *follows* the gross shape of the target reflectance profile, at least to the degree possible)”

(Churchland, 2007, p. 208, italics original).

Relevant for our purposes is Churchland’s proposal that *the tilt-angle* of these resultant ellipses be identified with the saturation of a colored surface (212). The argument for the impossibility of a supersaturated orange reflectance (e.g.), then, proceeds as follows:

The highest possible (i.e., the maximal) saturation for orange reduces to a particular tilt-angle,  $n^\circ$ , which is actually instantiated (*ibid.*, p. 189). You need only stare for a prolonged period at a richly saturated pale blue-green surface before looking at an orange surface which instantiates a tilt-angle of  $n^\circ$  in order to induce an experience as of an orange *more saturated* than that maximally saturated orange. The resultant experience is as of an impossibly saturated orange surface.

This quick demonstration relies on the assumption that saturation reduces (to tilt-angle). But what if we think, instead, that it does not reduce to tilt-angle but merely supervenes on it. In that case, we need a revised argument for the impossibility of supersaturated orange. But it is not far from the case just made. That argument proceeds like so:

There is no tilt-angle  $>n^\circ$  compatible with the objects of color experience being CA ellipses. So, there is no tilt-angle  $>n^\circ$  which instantiates (underpins) a level of saturation beyond that instantiated by  $n^\circ$ . It would have to be a property distinct from tilt-angle which instantiates that level of saturation. But if saturation is ever instantiated by

properties distinct from tilt-angle, then it is no longer true that saturation supervenes on tilt-angle. This gives us a contradiction. Therefore, if saturation supervenes on tilt-angle, then there is no level of saturation greater than that level instantiated by tilt-angle  $n^\circ$ .

# Russellian representationalism and the stygian hues

## **Abstract:**

Representationalism, which says that phenomenal character supervenes on representational content, is today the leading theory of the phenomenal character of perceptual experience. And Russellian representationalism (RR), which specifies that experiences contents are Russellian propositions/object-property pairs, is the leading iteration of that theory. In this paper, I show that the stygian color experiences (Churchland, 2005) serve as counterexamples to RR.

Phenomenally distinct stygian color experiences have the same Russellian contents. Hence, the experiences serve as counterexamples to RR.

Representationalism is today the leading physicalist theory of the phenomenal character of perceptual experience. And Russellian representationalism, which identifies contents with extensions, is the leading iteration of that theory. If there exist phenomenally distinct experiences as of the impossible, then these would *prima facie* serve as counterexamples to the theory. In order that they definitively serve as counterexamples, it needs to be that there is no plausible account of the experiences on which they decompose into constituent elements each of which is unproblematic from the perspective of the theory. The contention of this paper is that the *stygian color experiences*, afterimage-experiences as of maximally dark, hued surfaces, of Churchland (2005) serve as counterexamples to Russellian representationalism.

The plan for this paper is, first, in section 1, to get clear on exactly what metaphysically Russellian representationalism is committed to and what motivates accepting it over its rivals. In section 2, we will see why phenomenally distinct experiences as of the impossible serve as counterexamples to the theory barring a plausible decomposition of the associated contents into possible constituents. In section 3, we introduce the stygian color experiences, which, it will be argued, are experiences as of the impossible (section 4) which do not plausibly decompose into possible constituents (section 5). I respond to objections in section 6 before concluding.

### *1 The target*

Representationalism is usually spelled out in terms of supervenience: as saying perceptual phenomenal character supervenes on representational content. This definition excludes “weak” versions of representationalism, which say only that phenomenal character has attendant content but not that it supervenes on content.

Supervenience is most easily understood as a claim about when changes in one type of phenomenon are possible, namely, only when there is a change in some other type of phenomenon. There can be no change in phenomena of type *A*, it says, without a change in phenomena of type *B*. But it might also be explicated in terms of duplication: whenever objects

or properties of type *A* are duplicated, so too are the objects or properties of type *B*. We will make use of both construals of supervenience below.

The representational content of a perceptual episode is whatever the episode “says” about the world outside—whatever it says of the subject’s environment (or sometimes body). Conjoining supervenience and content thus characterized, we get that representationalism says there can be no change in the phenomenal character of a perceptual episode without a corresponding change in the episode’s content. This is representationalism about perceptual consciousness.

There is a long-running question among representationalists about what metaphysical gloss to give to content. Is the content of an experience a Fregean sense, these being intensions or modes of presentation? Or is it a Russellian content, a structured object-property pair (or structured proposition involving objects and properties)? Or maybe it is the set of worlds wherein the experience accurately portrays things. The most popular option today is to identify experiential content with Russellian propositions/object-property pairs.<sup>56</sup> (For ease of illustration, I will speak in terms of object-property pairs. Nothing is lost in doing this.)

One might prefer this view of what content is because one is antecedently attracted to Russellian contents in the case of propositional attitude content. In the case that one is, opting for non-Russellian contents, like sets of possible worlds or Fregean senses, in the case of experience would be to multiply entities needlessly if Russellian contents would do the job just as well. Moreover, insofar as one is antecedently unattracted to non-Russellian contents, she will find non-Russellian versions of representationalism unattractive.

---

<sup>56</sup> I follow Chalmers (2010 pp 356-61) in classifying Maund (1995), Holman (2002), Jakab (2003), and Wright (2003) and Thau (2002) as at least compatible with Russellian representationalists. Today, we can add to that list Tye (2009) and Speaks (2015).

Possible worlds representationalism has fallen out of favor in the last couple of decades. It is the Fregean version of the theory whose popularity rivals that of the Russellian version. The Russellian story about contents is simpler than the Fregean one: the Russellian account posits just extensions as contents while the Fregean account posits extensions and intensions (cf. Chalmers, 2010 p 361). So, if the complexity of the Fregean story of content can be avoided, then the representationalist should posit the simpler Russellian contents.

Committing to content's identity with Russellian contents gives us the following strand of representationalism about perceptual consciousness:

**Russellian representationalism** =<sub>def</sub> perceptual phenomenal character supervenes on perceptual experiences' Russellian contents.

On this view, when a subject has a visual experience of a red ball before her, the episode's content we can write like so: [that ball, red]. And when she has an experience of an orange cube, we write it [that cube, orange]; and so on.

### *1.1 The constituents of Russellian contents*

A number of considerations over the years have pushed proponents of RR to modify the object-involving portion of the content—of the lefthand side of our transcription of the content. Products off an assembly line, for instance, being qualitatively identical, will induce phenomenally identical experiences in subjects but involve different objects. If Russellian contents are object-involving, then the contents of a series of experiences of the numerically distinct products will be the same in each case. Three visually experienced red balls off an assembly line will induce content [that<sub>1</sub> ball, red], [that<sub>2</sub> ball, red], and [that<sub>3</sub> ball, red], where the demonstratives' subscripts indicate that distinct demonstrative-tokens are used in each case and, so, that different balls are involved in each case. These contents are all different. This has compelled some proponents of RR to replace *that ball* above with an existential generalization

with a location-property attached: [there is some ball, red & such-and-such location] (cf. Chalmers's discussion, 2010, p. 358).

But the issue of how best to characterize the object-side of Russellian object-property pairs should not detain us because the problem I will be raising for the theory regards just the property-side (the righthand side) of the content.<sup>57</sup> What matters for us is the following question: What kind of property is *red* in contents like [that ball, red]? And all versions of RR agree on this much: what bracketed 'red' signifies is a property, and the Russellian object-property pair is literally constituted by the property experience attributes to the ball.

## *2 Experience as of the impossible as a problem for RR*

One sort of counterexample to RR comes from experience as of the impossible. Take, for example, a visual experience of Escher's famous ever-ascending staircase that connects in on itself. Let us call the property of being so arranged *being an Escher staircase*. Because the property *being an Escher staircase* is impossible and, so, exists in no world, the corresponding Russellian content for such an experience would be [ $\emptyset$ , ...].<sup>58</sup>

---

<sup>57</sup> Though perhaps it deserves comment that representationalism does not predict what a change in the content of an experience should result in. It predicts only what a change in phenomenal character should result in (namely, a change in content). The two consequences we have just seen of phenomenal character's supervenience on content are that there will be no change in phenomenal character without a change in content and that content-duplicates are phenomenal duplicates. That there should be no change in content without a change in phenomenal character is an unrelated thesis. So, that numerically distinct experiences have different contents does not bear on what the experiences' phenomenal character should be. Accordingly, if there is any advantage to making Russellian contents object-involving, then content-distinct, phenomenally-identical experiences should not deter us from making them object-involving.

<sup>58</sup> I use the ellipsis here to signify that a constituent is lacking in the righthand position of the object-property pairing. Later, where context makes it obvious, I will use ellipses to elide members of a series.

This generalizes. If any experience predicates of some object,  $o$ , an impossible property, the associated Russellian content of the experience will be  $[o, \dots]$ . Why? For now, we will say that this is because if the property exists in no world, then it does not exist full-stop and, so, cannot constitute the Russellian content. But we will come back to this issue in the second of the Objections below (section 5).

As it happens, it is no problem for RR if there is only one type of experience which predicates an impossible property of  $o$ . It is bad if phenomenally distinct experiences do this. This is because if we have two or more phenomenally distinct experiences each with content  $[o, \dots]$ , then we have sameness of content but distinctness of phenomenology, which our definition of RR explicitly disallows. So, we cannot have two or more phenomenally distinct experiences each without a corresponding property.

And there *are* other experiences as of impossible scenes. Take, for instance, the waterfall illusion, which is a visual experience as of a moving-yet-stationary waterfall. Granting that a moving-yet-stationary waterfall is impossible,<sup>59</sup> the Russellian content of such an experience would also be  $[o, \dots]$ .

What it's like an undergo the waterfall illusion is distinct from what it's like to see an Escher scene. But the Russellian contents of each experience will be  $[o, \dots]$ . The examples, then, give us phenomenal distinctness in the absence of distinctness of content, and, so, *prima facie* serve as counterexamples to RR.

But these are familiar cases that have already been addressed in the literature, and I will be assuming that they can be handled in the following way. Michael Tye, in his (2000), suggests that experiences as of impossible scenes be handled by analyzing their contents, not wholesale, but in

---

<sup>59</sup> If we do not think the example works, there is a plethora of other shape-examples from which to choose, like the Penrose triangle. And there are even aural examples, like the (aural) barber pole illusion.



a way which respects the content's generation by distinct information channels in the visual pathway. He writes,

“Given the complexity of the content of visual experience and the number of different channels of information that lie behind its generation, it should not be surprising that in some cases an overall content is produced that is internally inconsistent” (2000, p. 75).

The idea seems to be that the proprietary informational contents of the distinct information-channels implicated in visual information processing will feature directly in the experience-content. And sometimes, as in visual experience as of impossible scenes, these distinct informational contributions come together in such a way that the overall experience-content is contradictory or inconsistent.

Applying this strategy to visual experience of an Escher scene, the idea seems to be that we can say the following. Distinct information-channels implicated in the visual pathway provide, respectively, informational contents  $I_1, \dots, I_n$ , where each  $I_k$  is internally consistent. But the conjunction of some or all of the  $I_i$  is inconsistent or contradictory. Moreover, and crucially, because each of the  $I_i$  is possible, there is no risk of the relevant Russellian contents' lacking suitable constituents. So, the content of the Escher experience is  $[0, I_1, \dots, I_n]$ , where there are no gaps to worry about. And the content of a phenomenally distinct experience as of an impossible scene, like the waterfall illusion, will have distinct constituents,  $[0, I^*_1, \dots, I^*_n]$  where again there are no gaps to worry about. Phenomenally distinct experiences, then, are shown to have distinct contents, and the counterexamples are avoided.

Now, if we had not had these sorts of decomposition stories available to tell, then experience as of impossible scenes, like the two just discussed, would already have successfully served as counterexamples to RR. If RR is a theory worth arguing against anew, then we need to assume that these sorts of decomposition stories are available. What we should be saying, then, is this: phenomenally distinct experience as of the impossible will only serve as counterexamples to RR

if no plausible decomposition story, like the ones just discussed, is available.<sup>60</sup> If we have experiences as of the impossible which do not plausibly decompose, then we have a counterexample to RR. In the next section, I will introduce experiences which do just that.

### *3 Churchland's stygian color experiences*

We have just seen that phenomenally distinct experiences that are each as of the impossible would serve as counterexamples to RR unless the experiences' associated Russellian contents plausibly decompose into constituent elements each of which is possible. RR would assign them, in spite of their phenomenal distinctness, identical contents. And RR explicitly disallows such things.

And the contention of this paper is that the *stygian color experiences*, afterimage-experiences as of impossibly dark and yet hued surfaces, of Churchland (2005) give us just this sort of counterexample. RR assigns each phenomenally distinct member of this class of experiences identical contents because each phenomenally distinct such member is a case of experience as of *an impossible color property*. (This will be argued for in section 4.)

And, so, take as two examples stygian yellow and stygian blue experience. These are afterimage-experiences as of something dark-as-black and yet distinctly yellow and as of something dark-as-black and yet distinctly blue. No color property corresponds to either experience type—each is an experience as of an impossible color property.

---

<sup>60</sup> As an anonymous reviewer pointed out, even in the example cases given, we might think the decomposition leaves something out: there is a difference between experience of the parts of an impossible scene and experience of the whole scene. We can speculate what sorts of things Tye might say in reply. Perhaps the latter is different for giving rise to a judgement “that cannot be!” which judgment has a specific cognitive phenomenology. I will leave the worry here, though, because I am assuming for argument's sake that the strategy is successful.

Inducing these afterimages in your own experience is easy.<sup>61</sup> To induce, e.g., a stygian blue experience, fixate for approximately 20 seconds on the black crosshair in the yellow circle on the top left of **Figure 1**. Immediately afterwards, transfer your gaze to the first black surface to the right, and you should have the experience: an afterimage-experience as of something dark-as-black and yet distinctly bluish. In order to induce stygian yellow experience, fixate for approximately 20 seconds on the black crosshair in the blue circle of the fourth row of **Figure 1**. Immediately thereafter, transfer your gaze to the first black surface to the right and you should have the experience: an afterimage-experience as of something dark-as-black and yet distinctly yellowish. (The rightmost column contains, as it says, “rough predictions” of what your experience will be like. These can only be rough for reasons that will become apparent in section 4.) To induce stygian green and stygian red experience, follow these same instructions on rows 3 and 4. These are the afterimage-experiences that I will be arguing serve as counterexamples to RR. Each is an experience as of an impossible color property.

---

<sup>61</sup> Which is to say nothing of the ingenuity it took to predict the experiences.

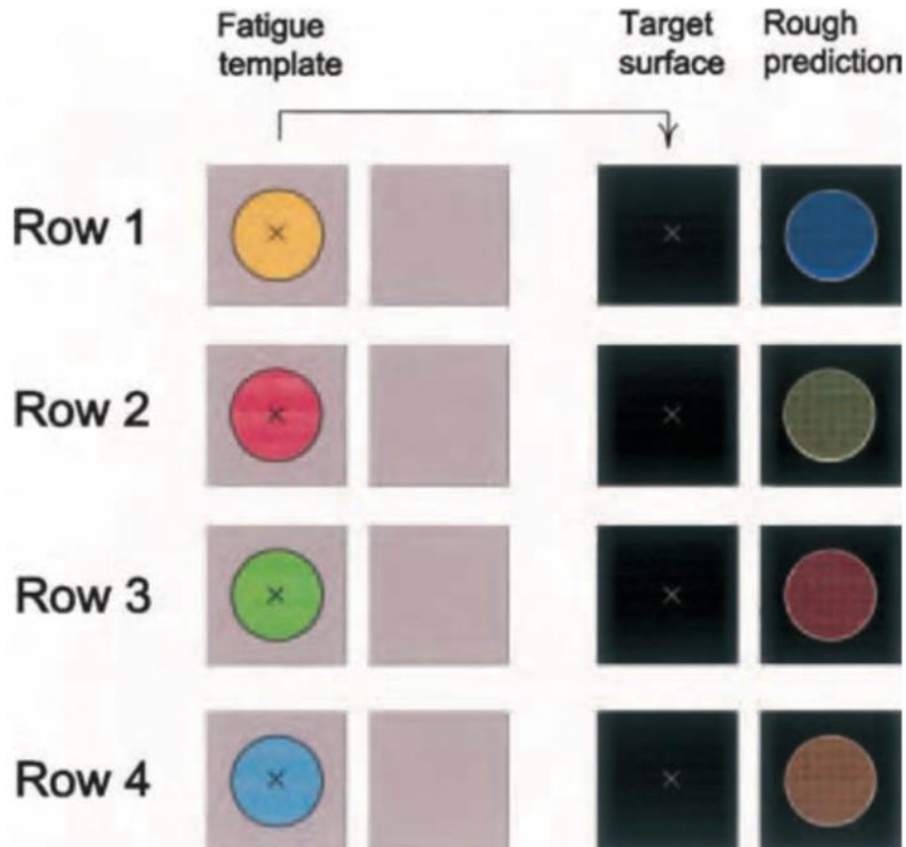


Figure 1. Aid to produce stygian color experiences. Reproduced with permission from Cambridge University Press.

It is important we know what the neural mechanism is which underpins these afterimages because its operation will be directly relevant to the case to be made in section 5 that no plausible story is available to my opponent on which the stygian color experiences decompose into constituent elements each of which is possible. But a just cursory glance at the relevant science should do. (The reader already familiar with the Hurvich-Jameson opponent-process theory of color experience is advised to skip to the next section, section 4, where the case will be made that RR assigns phenomenally distinct stygian color experiences identical contents.) Knowing just the

following five details of the Hurvich-Jameson opponent-process theory of human color experience will suffice for our purposes.<sup>62</sup>

1) The stygian color experiences are neurally underpinned by the joint activity of three cell-types that are found downstream from retina but early on in the visual pathway: the so-called Blue/Yellow (B/Y), Green/Red (G/R), and Black/White (B/W) color-opponent neurons, each of which can undergo the full range of levels of activation from 0% - 100%, with the default resting state of each cell-type being 50% of full activation (Churchland, 2005, pp. 164-65). 2) The full variety of joint activation-levels of the B/Y and G/R opponent-neurons underpins all experience of *hue* and *saturation*, and the full range of activations of the B/W cells underpins experience of hue *lightness/darkness* (165-66). See **Figure 2** for a diagram of the entire range of human color experiences which result from the full variety of the above three cell-types' joint activations. 3) The stygian color experiences are all underpinned by i) maximal inhibition of Black/White opponent-cells and ii) *some* inhibition or excitation of the former two cell-types (179-85). (The former two cell-types cannot be left at their default 50% activation because the satisfaction of (i) but not (ii) underpins experience of the achromatic colors: white, black, and the scales of gray.) 4) It is three retinal cell-types, S, M, and L cones, whose activations affect opponent-neurons, and they respond differentially only to short (S), medium (M), and long (L) wavelength light, with one cell-type designated to each length (*ibid.*, p. 164). And lastly 5) The G/R neurons get inhibited exclusively by M cones and get excited exclusively by L cones; the B/Y neurons get inhibited exclusively by S cones and get excited by both M and L cones but not S cones; and the B/W neurons respond to cones which themselves are sensitive to the amount of light, of any wavelength, impinging on the S, M, and L cones in their vicinity (164).

---

<sup>62</sup> For a thorough presentation of the relevant science, see Hardin (1988) or Churchland (2005). For details, I rely on the latter presentation.

This concludes our very brief look at the opponent-neural machinery which underpins human color experience, including, relevantly for our purposes, the stygian color experiences. We can now move on to the case for the impossibility of the color properties these experiences represent.

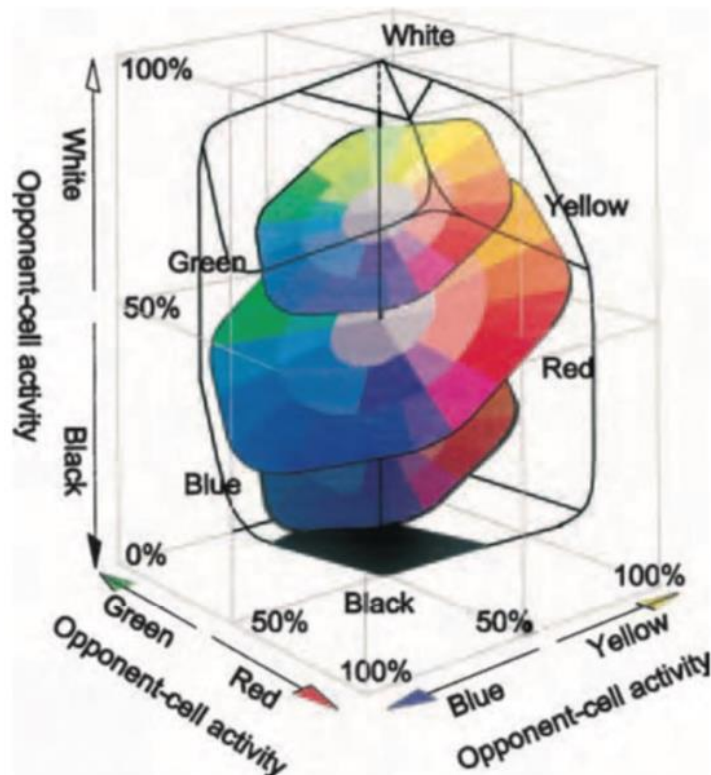


Figure 2. Range of color experiences and their corresponding activations. Reproduced with permission of Cambridge University Press.

#### *4 The stygian color experiences are experiences as of the impossible*

In the previous section, we saw how to induce the stygian color experiences in ourselves. We saw that by fixating on colored surfaces for a prolonged period and subsequently fixating on a black surface, we can undergo afterimage-experiences that are maximally dark and yet distinctly hued. These uncanny afterimage-experiences bear on RR because each of them is an experience as of an impossible color property: no object can instantiate maximal darkness and yet also be hued.

RR, then, assigns the experiences identical contents—in each case, [*o*, ...]—despite their phenomenal distinctness.<sup>63</sup>

In the following subsection, 4.1, I will argue that there is no reflectance property with which the stygian colors are identical. And in 4.2, I will argue that neither are there any primitive color properties with which the colors are identical. This will show that the stygian hues are *metaphysically impossible* qua reflectance property and qua primitive color property. (This leaves open that the stygian colors are possible qua dispositional or relational color properties. I will address the stygian colors' possibility qua dispositional or relational color properties in the Objections and Replies.)<sup>64</sup> Note that these are both physicalist theories of color. I restrict my attention to physicalist theories of color in this way because RR is a physicalist account of

---

<sup>63</sup> I say “uncanny”. But it may just be that they seem uncanny when we place them in the context of expectations we have for colors in broad daylight, for instance. The afterimages may not feel so different from, say, phosphene experiences, which are not unordinary. Note, though, that it only bodes worse for RR if the afterimages do strike us as ordinary, because then RR fails to accommodate experiences that are ordinary.

<sup>64</sup> Churchland provides his own arguments for the stygian color properties' impossibility in the (2005) paper (pp. 182-83). But it suits his aims in that paper to provide what are comparatively informal arguments. Churchland's principal aim there is to show that, in a modest but significant way, physicalists can give their opponents a case of physical facts entailing phenomenal ones. As will be familiar, dualists (like Jackson, 1982, and Chalmers, 2010) argue that a necessary condition on physicalism's truth is that the facts of physics entail the facts about phenomenal character. Moreover, the stygian color experiences' import lies, for Churchland, in their *novelty*. A hallmark of a scientific theory's demonstrated success is the panning out of its novel predictions. And, so, it is not crucial for Churchland's purposes that the stygian hues be genuinely impossible. He only needs that they be the sorts of things one should be surprised to learn of. If they are surprising in this way (for intuitively seeming impossible, we can grant), then the Hurvich-Jameson model's predictive power becomes even more remarkable. It is important for him to make the case for the stygian hues' impossibility only insofar as this underscores their novelty. Accordingly, I need to say more to make the case that stygian color properties are metaphysically impossible and that they are metaphysically impossible *qua* reflectance and primitive color property.

perceptual phenomenal character. The contents on which experience supervenes, then, *prima facie* should be physically constituted too, lest RR lose its claim to physicalism. This is a connection, however, that might be contested. Accordingly, I will consider in the Objections and Replies what the costs are of slackening commitment to the exclusively physical constitution of content.

#### 4.1 Reductionism

On reductionism, the colors are identified with *reflectances*,<sup>65</sup> a reflectance being a measure of how much of the different parts of the visible portion of the electromagnetic spectrum a surface reflects.<sup>66</sup> There are two intuitive arguments for the stygian hues' impossibility (qua reflectance), which, when taken together, mount a strong case against their possibility. The first says that, given what we should already be saying about when a surface qualifies as maximally dark, or hued, we should want to rule out the possibility of the stygian reflectances on these grounds. The second says that any proposed reduction of the stygian colors would be exceedingly ad hoc and, so, should be avoided.

First, on a metaphysics of color which identifies colors with reflectances, arguably it is legitimate to call an object *dark as black* ("maximally dark") insofar as its reflectance tends towards being flat up against the wavelength axis of its associated graph (cf. Tye, 2000, p. 157; Churchland,

---

<sup>65</sup> Or something thereabouts: Byrne and Hilbert (2003) identify the colors with disjunctions of general tendencies to send light thus-and-so ("send" here being a catch-all for reflectance, refractance, emittance, and the rest). And Churchland (2007) identifies colors with canonical approximation ellipses, a mathematical feature of reflectances that each colors' metamers share (metamers, roughly, being differences in reflectances that subjects cannot detect).

<sup>66</sup> There are more ways to be colored than to have a reflective surface—many colored objects do not reflect light but transmit it, refract it, emit it, etc. (see previous note). But it will simplify discussion to consider just reflectances.

Invoking the other manners by which an object might send light thus-and-so does not help my opponent, for there is no way at all to be dark-as-black and hued.



2007, p. 212). The more an object's curve strays from that axis, the more legitimate it becomes to call the object *hued*. And, so, it will be legitimate to call an object both *hued* and *dark-as-black* insofar as its reflectance is simultaneously flat up against, and not flat up against, the wavelength axis of its associated graph. I.e., it will never be legitimate to do so. So, it will never be legitimate to call any possible reflectance a *stygian color*.

Second, the only other candidate reflectances involve reference to portions of the electromagnetic spectrum *outside* the visible range. But identifying the *stygian colors* with reflectance defined in terms of more of the EM spectrum than the visible range—in terms inclusive of, say, ultraviolet or infrared—would be exceedingly arbitrary, unmotivated, and *ad hoc*.<sup>67</sup> We should accordingly avoid positing of any reflectance defined in terms of the EM spectrum beyond the visible portion that it is identical with a *stygian color*.

The upshot is that it is better to think of *stygian reflectances* as impossible rather than possible. This because, first, their impossibility is suggested by what the reductionist is already advised to say regarding what qualifies as *black* and what qualifies as *hued*. And second, because any proposed reduction, one, say, inclusive of more of the electromagnetic spectrum than the visible range, is bound to be exceedingly *ad hoc*.

#### 4.2 *Primitivism*

The *stygian colors* are also impossible *qua* primitive color properties. On primitivism, the colors are said to be irreducible physical properties of objects.<sup>68</sup> On this view, the colors are not

---

<sup>67</sup> It certainly is conceivable (in the sense of conceptually non-contradictory) that *stygian yellow* is instantiated by objects which reflect radio waves *thus* and the visible portion of the spectrum *so*. But to call *stygian yellow* possible in light of that is a mistake because it is equally conceivable that *redness* is identical to that same property. This sort of conceivability is not a guide to possibility—at least not in the present context.

<sup>68</sup> The locus classicus here being John Campbell's (1993) paper "A Simple View of Colour".

identical to paradigmatically physical properties—to properties of the so-called scientific image. And yet they are physical. They are physical because they *supervene* on properties of the scientific image—on reflectances, in particular (Byrne and Hilbert, 2006, p. 75). Further, the colors here are not individuated with reference to subjects or subjects' responses. They have their natures independently of their disposition to affect subjects like us in the ways that they do. Finally, on this view the colors are primitive: that is, they are not reducible to any more fundamental physical properties.

Let us work with the example of stygian yellow. Is *primitive stygian yellow* instantiated in any world? We cannot say here, as we did in the context of reductionism, that something is black insofar as its (associated) reflectance tends towards being flat against the wavelength axis of its reflectance profile. Blackness may supervene on flat such curves, but these physical goings on are not a part of blackness's nature. Similarly, we cannot say that something is hued insofar as its associated curve strays from the wavelength axis of its reflectance graph. Colors' natures are distinct from such physical properties. Nevertheless, unpacking primitivism's supervenience claim will allow us to see why primitive stygian colors should not be thought to be possible, either.

In Byrne and Hilbert's seminal (2006) discussion, we see that the primitivist's supervenience claim is spelled out in one of three ways in the literature. On the view, colors are *nomologically coextensive with*, *metaphysically determined by*, or *metaphysically coextensive with* reflectances. On the first construal of the supervenience claim:

NC) "For any color  $c$ , there is a [reflectance]  $P$  such that  $P$  is nomologically coextensive with  $c$ . Equivalently: it is a law that for every object  $x$ ,  $x$  has  $P$  iff  $x$  has  $c$ " (75).

On the second:

MD) “For any color  $c$ , there is a [reflectance]  $P$  such that  $P$  metaphysically necessitates  $c$ .

Equivalently: it is metaphysically necessary that for every object  $x$ , if  $x$  has  $P$ ,  $x$  has  $c$ ”

(*ibid.*).

And on the third:

MC) “Colors are (metaphysically) necessarily coextensive with [reflectances]” (76).

How do these theses bear on the possibility of the stygian hues? Well, these formulations tie the colors tightly to their supervenience bases. On NC, in order for a primitive stygian color to be instantiated in a world it needs a reflectance to instantiate it. This is a consequence of the thesis’s ‘iff’ connective. The third fleshing out of the supervenience claim, MC, also makes it so that primitive stygian colors will not be instantiated absent some reflectance or other. This is a consequence of the coextension of colors and reflectances. If a color were instantiable in the absence of an underlying reflectance, colors and reflectances would not be coextensive.

The problems with locating primitive stygian yellow in a world if NC or MD is the case are strictly analogous to the case made in the context of reductionism. First, we said above that we should not think of there as being any reflectances associated with the stygian colors. Because there are no reflectances, neither could there be primitive stygian properties which supervene on those reflectances. And second, positing of any reflectance that stygian yellow is instantiated exactly when it is instantiated is as ad hoc here as it was in the context of reductionism.<sup>69</sup>

However, on the second fleshing out of the primitivist supervenience claim, MD, though reflectances’ instantiations will suffice for the instantiation of various colors, the instantiation of primitive stygian colors is not so tightly tied to reflectances. Every world with a particular reflectance has a particular color, yes; but it is not the case that every world with a particular

---

<sup>69</sup> And as before (fn. 67), conceivability is not a guide to possibility here. It is just as conceivable that redness supervene on whatever property conceivably subvenes primitive stygian yellow.

color has a particular reflectance. This gives the primitivist proponent of RR room to hazard that primitive stygian yellow (e.g.) is instantiated by *some property altogether distinct in kind* from reflectances. Can the primitivist exploit this looser connection between color and supervenience base to accommodate primitive stygian colors' possibility?

By my lights, there are three ways she might attempt to exploit this looser connection. She could posit primitive stygian colors' supervenience i) on actual properties which are not reflectances, ii) on non-actual properties which are not reflectances, or iii) on alien properties (which are, perforce, non reflectances). I will address what the problems are with each of strategy in turn.

Here is why the primitivist should not posit primitive stygian colors' supervenience on actual properties which are not reflectances. Take as an example positing that primitive stygian yellow supervenes on massive quarks' decaying into less massive quarks. Hazarding supervenience bases this exotic is ill-advised. If we posit exotic supervenience bases like this, we need to say that in every world that this base property is instantiated so is (primitive) stygian yellow. And no one should think that in the actual world, stygian yellow is instantiated whenever and wherever quarks decay. The example was intended obviously to be strange, but the point generalizes. As before, we should not say that any physical property in the actual world instantiates stygian yellow because we have no reason to think that any property in the actual world correlates with stygian yellow. The general lesson is that the danger of saying primitive stygian yellow is instantiated in any world,  $w$ , is that we (on MD) are thereby saddled with positing its instantiation in the actual world wherever the property which underlies its occurrence in  $w$  is instantiated in the actual world.

Let us turn to option (ii). Could the primitivist proponent of RR anchor primitive stygian yellow to physical properties that do not exist in the actual world—like, say, the property of being a sphere of gold 100m in diameter?

Doing that is also unmotivated and ad hoc. If one anchors primitive stygian yellow to the property of being a gold sphere 100m in diameter, then one is stuck with saying, given the relevant supervenience commitment, that were the property of being a gold sphere 100m in diameter to be instantiated in the actual world, the sphere would be (primitive) stygian yellow. We shouldn't say this. Further, one would be stuck with saying that in all worlds where the property of being a gold sphere 100m in diameter is instantiated, primitive stygian yellow is too. This would be a fact about all those worlds. We could only pretend at having reason to think that the giant, gold spheres of those worlds really are stygian yellow.

And finally, let us consider option (iii). What is the problem with positing primitive stygian colors' supervenience on alien properties. An alien property is one that neither is instantiated in the actual world nor is a combination of any actually instantiated properties. I have no reason to deny that primitive stygian yellow supervenes on alien properties in remote alien worlds. But there is the following problem for this move: alien properties are not physical. Why not? On at least one influential definition of the physical, a property is physical only if it is referred to in fundamental physical theory (Stoljar, 2010, p. 57)—or, I think we should also add, if it is realized by the states/properties referred to by those theories. And alien properties are neither referred to in fundamental physical theory nor realized by the states/properties referred to therein. (We moreover should not anticipate that future fundamental physics advantage itself of properties which neither exist anywhere in the history of the actual world nor might be made by combination of any such properties.) But we call primitive color properties physical because they supervene on physical properties. Primitive stygian colors which supervene on nonphysical alien properties lose their claim to physicality. And this bodes badly for RR because RR is a brand of physicalism. Its proponents, accordingly, will want every perceptual experience to be physical and, so, supervene on the physical. So, the proponent of RR should not have primitive stygian color properties supervene on alien properties.

The verdict, then, is this. We have no reason to say that primitive stygian colors are instantiated in any world. When we say they are instantiated in a world wherever there is physical property P, we are thereby committed, via the supervenience claim of primitivism, to saying this same thing about the actual world: wherever P occurs actually, primitive stygian colors do too.<sup>70</sup> Nothing motivates locating primitive stygian colors in the actual world; any such locating of them in actuality is ad hoc. Moreover, nothing motivates anchoring primitive stygian colors to physical properties that do not exist actually, like the property of being a gold sphere 100m in diameter, nor on alien properties. To do so would be to make ad hoc claims about swathes of modal space, or it would be to give up on physicalism. We would do better, then, to hold that primitive stygian colors are impossible.

This concludes the argument that the stygian colors, qua property of external object, are impossible. Demonstrating their impossibility gets us halfway to the stygian color experiences' serving as counterexamples to RR. What needs finally to be shown is that no plausible story is available to the proponent of RR whereunder the stygian color experiences decompose into possible constituents.

##### *5 The stygian color experiences decompose at too steep a price*

We have said that phenomenally distinct experiences serve as counterexample to RR only if no plausible story can be told whereunder the experiences' contents decompose into constituent elements each of which is possible. And we just saw that the stygian color experiences are as of

---

<sup>70</sup> Cf. also Macpherson (2003): "on an objective physicalist theory, once we have singled out the physical properties that in our world are responsible for colour, those physical properties are the colour properties in all possible worlds. Colour words are taken to refer rigidly to the physical properties so identified....It is crucial to the objectivity of the theory that colour words rigidly refer in this way and that the logical independence of colour properties from colour experiences is maintained" (54-55).

impossible properties. What remains to be seen, then, is that these contents do not plausibly decompose.

Crucially, two things need to be in place in order that experiences as of the impossible plausibly decompose into possible constituent elements. 1) It must be that there really are distinct information channels responsible for subjects' representing the prized apart content-components. And 2) the properties about which these channels carry information must be possible.

With this in mind, the most natural decomposition of the stygian color experiences treats their contents as being constituted by distinct hue and lightness/darkness components. The most natural decomposition of, e.g., stygian yellow experience sees its content as being constituted by distinct *yellowness* and *maximal-darkness* components: [*o*, yellow & maximally dark].

Opting for this treatment plainly satisfies condition (2): both yellowness and maximal-darkness are possible, as evidenced by the existence of yellow things and black things. What about condition (1)? According to the opponent-process theory of color-experience that we are relying on (section 3), there are indeed distinct information channels involved in stygian color experience (involved in *any* color experience). Particularly, it is one type of cell, the B/W opponent-neurons, which carries information about the lightness/darkness of a hue. And it is two other cell-types, B/Y and G/R opponent-neurons, responsible for information about hue (and saturation). So, the current strategy is at least *prima facie* available to the proponent of RR.

But the problem with this strategy, to put it provocatively, is that if we go all the way with it, *then color experience ceases to be about color*. Less provocatively: If we want to think of each information-channel that has a distinctive informational-profile as contributing a proprietary informational content to the Russellian content of an experience, then the channels which underpin experience of hue (and saturation), these being, recall, the G/R and B/Y opponent-cells, must be prized apart as well. This because they too have distinct informational-profiles, neither of which

includes reference to the colors (more on why in the next paragraph). So, once we do think of every channel with its own distinctive informational-profile as contributing different types of content, then the colors no longer feature as constituents in the contents of color experience. And the only thing which could motivate keeping these channels' contributions joint would be a desire to avoid the counterexample. This is to say that experience-content will inevitably be unacceptably gerrymandered if it is to be constituted by hue and lightness rather than all of the informational contributions which would feature in the content were we to fully pursue the current strategy.

To see that the information-channels which underpin registration of hue (and saturation) do not carry information about the colors, consider just the channel constituted by G/R opponent cells. We saw above that this channel carries information about net differences in long wavelength light versus medium wavelength light impinging on retina (Churchland, 2005, p. 165; see also Tye, 2000, pp. 160-61).<sup>71</sup> But the colors are not identical, nor supervenient on, differences in the long and medium wavelength light that impinges on retina. The colors, we have said, are reflectances, or properties which supervene on reflectances. Moreover, very many things which are not green or red can send *that* light to the eye: blue objects behind yellow films, white objects bathed in green light, white objects bathed in red light, and so on and so forth. So, if we think of the G/R channel as contributing information about these differences in incident light, then the colors no longer feature in color-experiences' Russellian contents. And being forced to take color out of color experience, I will assume, is too considerable a cost for the representationalist to incur.

---

<sup>71</sup> The information carried by B/Y cells is slightly more complicated, involving further computations over signals sent from retina (Churchland, 2005, p. 165) . But it is true of B/Y cells too that color is not what they inform about. It is the joint operation of the three cells-types—or at least of the G/R and B/Y neurons—that carry information about anything recognizably color-like.



This response raises the obvious question, Then what *does* carry information about color in the brain? Presumably, the neural activity which carries information about the colors is that activity which underpins color-constancy. And color-constancy cannot be achieved in abstraction from registration of lightness/darkness. That is, constancy is not exhibited absent B/W opponent-cell activity. But once we yoke the informational contributions of B/W opponent-neural activity to the informational contributions of the activity of the cells responsible for experience of hue, we reintroduce stygian color properties into our contents. We are back to assigning identical contents to phenomenally distinct experiences.

The upshot, then, is that we do not have a plausible story to tell whereunder the stygian color experiences decompose into constituent elements each of which is possible. And this lack of plausible decomposition story, in the face of what we established in the previous section, namely, that stygian color experiences give us experiences as of impossible properties, implies that the stygian color experiences serve as counterexamples to RR.<sup>72</sup>

### *5.1 Recap*

Before turning to objections, let us briefly recap how the argument of the previous sections comes together. We have a counterexample to RR when we have phenomenally distinct experiences with identical contents. The stygian color experiences, stygian yellow and stygian blue, for instance, are phenomenally distinct experiences with identical contents. These contents

---

<sup>72</sup> There is now enough in place to explain how my counterexample differs from Macpherson's (2003) example of experience as of red-green. In cognitive neuroscientific work that post-dates Macpherson's article, experience of red-green has been taken to evidence that the hue-solid has distinct dimensions responsible for red and green (Livitz et al., 2011). If that is right, then the proponent of RR could think of these dimensions as being underlain by distinct information-channels and accommodate Macpherson's counterexample that way. So long as the representationalist does not unhitch the B/W channel from the channels responsible for hue, she can prize channels apart like this without taking color out of color-experiences' Russellian contents.

are identical because they lack constituents in the same positions: a stygian yellow experience's associated Russellian content would be [ $\phi$ , ...], and a phenomenally distinct stygian blue experience's associated Russellian content would also be [ $\phi$ , ...]. And they lack constituents in these same places because, qua relevant property types (reflectances or primitive color properties), stygian yellow and stygian blue are impossible.

The stygian color experiences would not serve as counterexamples to RR if their associated Russellian contents plausibly decomposed into constituent elements each of which is possible. But the stygian color experiences' contents decompose at too steep a price. Their best hope for decomposition comes at the cost of color's ceasing to feature in color experience.

### *6 Objections and replies*

I want now to consider three objections. The first asks why the proponent of RR cannot slot metaphysically impossible stygian color properties into experiences' associated Russellian contents. The second asks why the proponent of RR cannot slot stygian color dispositions, or stygian color relations, into stygian color experiences' associated contents. And the third asks why the possibility of stygian colors cannot be accommodated by adopting color pluralism.

#### *Objection: Why not appeal to metaphysically impossible properties?*

A formerly popular version of representationalism held that experiential content is identical to sets of possible worlds (Tye, 1995; Lycan, 1996). Lycan, in addressing the issue of experience as of impossible scenes, says that the possible worlds representationalist should posit *impossible* worlds to accommodate these experiences (1996, p. 72). The question for us here, then, is whether the proponent of RR can make a similar move.

By my lights, there are at least three broad ways the Russellian representationalist might avail herself of a strategy in this vein. She could i) accept that metaphysical impossibilities exist and slot metaphysically impossible stygian color properties into the relevant contents; or ii) accept

that the stygian colors are metaphysically impossible but not logically impossible and slot merely logically possible stygian color properties into the relevant contents; or iii) accept that the stygian color properties are metaphysically impossible and posit their existence in an abstract realm and slot these abstracta into the relevant contents.

### *Reply*

Depending on other commitments of the Russellian representationalist, she may or may not be able to do this. Whether she can depends on whether she wants to *naturalize content*, in the manner of a teleosemantic or a causal theory of content. (And these two broad types, I take it, exhaust the externalist's prospects of naturalizing content.)<sup>73</sup> A word, then, on why these semantic-naturalization theses get in the way of employing any of strategies (i)-(iii).

To see why commitment to semantic-naturalization scuppers (i)-(iii), we can abstract away from the considerable differences between teleosemantic and non-teleosemantic causal theories of content. And this because they all have in common the following, quite broad, necessary condition:

In order for a state  $S$  of a system  $s$  to represent external object- or property-type  $F$ , some  $F$ s need to exist at some time in the past-present-future of  $s$ 's environment in order for the  $F$ s to causally interact with  $s$  or  $s$ 's ancestors.

Being metaphysically impossible, stygian color properties exist at no time in the past-present-future of any subject's environment. So, they cannot causally interact with any subject nor any subject's ancestors. Accordingly, neither a teleosemantic nor a non-teleosemantic causal story

---

<sup>73</sup> There are also structural resemblance/isomorphism accounts, as in Opie and O'Brien (2004), Churchland (2012), Ryder (2004), Shea (2018, ch. 5). However, with the exception of Opie and O'Brien (2004), these all have selectional or learning histories playing an indispensable role in content's grounding and, so, can count for our purposes as broadly teleosemantic.

could account for experience's representing stygian color properties.<sup>74</sup> That is, neither theory-type could accommodate the constitution of Russellian contents by stygian color properties.<sup>75</sup>

However, RR is not wedded to the project of naturalizing content. The proponent of RR could hold that, though content is natural, it is primitive. This would be to endorse a position in the spirit of Burge (2010). In that case, whether she can rely on the constitution of Russellian contents by metaphysically impossible properties depends on whether she is antecedently attracted to the existence of impossibilities. If she is not antecedently attracted to the existence of impossibilities, then relying on them to constitute the relevant contents is an ad hoc expanding of ontology. As a general rule, we should not bloat our metaphysics to accommodate our theories of mind. This, I take it, is a crucial element of the appeal of naturalist/physicalist approaches in philosophy of mind.

This last point goes for each of options (i)-(iii). If we are antecedently attracted to the existence of the impossible, the merely logically possible, or the merely abstractly existent, then we can have it that the stygian hues (qua reflectance or primitive color property) constitute Russellian contents.

This is all to say, then, that if we accept that phenomenal character supervenes on primitive content *and* that this primitive content is constituted by metaphysically impossible properties,

---

<sup>74</sup> Cf. again Macpherson (2003): “[E]ven if we were to allow that in a world with a physics very different to our own there could be colours that do not exist in this world, a teleologist or a causal covariation theorist would have difficulty establishing how experiences must relate to merely possible properties in order for those experiences to represent those properties” (p. 65 fn. 24).

<sup>75</sup> Note that this also blocks the proponent of RR from appealing to color eliminativism to accommodate the case, supposing she wants to commit to semantic naturalization. On color eliminativism, colors appear nowhere in the past-present-future of subjects' environments.

then we can avoid the problem posed by the stygian hues. But if the proponent of RR is not already attracted to such posits, then perhaps greener pastures lie elsewhere.

*Objection: Why not appeal to stygian color dispositions or stygian color relations?*

The second objection asks what is keeping us from thinking of stygian color experience as predicating stygian color dispositions, or stygian color relations, of the environment. On dispositionalism, the colors are identical to dispositions to elicit certain color-experiences (see, e.g., McDowell, 1985; Levin, 2000). If there is a world in which object surfaces are disposed to bring about stygian color experiences, then there is a world in which stygian color dispositions exist. The objections of section 4 do not count against worlds like those. So, I should grant the property is possible. It can, then, be thought of as an available constituent of the Russellian contents associated with stygian color experiences. And on relationalism, the colors are identical to triadic relations between subjects, objects, and viewing circumstances (see, e.g., Cohen, 2004), where the subject relatum, importantly, is a highly specified state of visual systems, and the viewing circumstances are highly specified as well (Cohen, 2009, p. 116). On relationalism, stygian colors would presumably be identical to relations between object surfaces, viewing circumstances which include in their specification the over-compensatory activity of implicated opponent neurons, and states of subjects' visual systems which include in their specifications the instantiation of states on the floor of the H-J space (neural states involving the full inhibition of opponent B/W cells). Such relations are instantiated every time subjects have stygian color experiences. Being *actual*, stygian color relations would be readily available constituents of the Russellian contents associated with stygian color experiences.

*Reply*

Let us start with stygian colors understood as dispositions. The role of experience in stygian color dispositions makes the dispositionalist metaphysics of color difficult to square with RR. The fundamental characterization of color given by dispositionalism is in terms of color-

experience. But the fundamental characterization of color-experience that RR gives is in terms of color. The circularity which results from conjoining these theses is, even if not vicious, at least too tight to be plausible (see Byrne, 2006, p. 225 and Levin, 2000, p. 164 for related complaints).<sup>76</sup> The proponent of RR as I have defined it should not posit the constitution of content by colors construed as dispositions.

Consider now stygian colors relations. This time, it is the role of subjects in stygian color relations that makes the relationalist metaphysics of color difficult to square with RR. In keeping with RR, experience as of stygian colors involves the representation of stygian colors. An initial worry we might have with the representation in experience of stygian color relations is that the neural vehicles of the representational contents (again, states on the floor of the H-J space) end up (partly) representing themselves.

To see this, take the vehicles of the contents whose constituents are stygian colors. These will be states on the floor of the H-J space some distance away from the resting points of the B/Y or G/R dimension. On the relational view, these vehicles themselves are partly constitutive of the stygian colors. So, the vehicles' contents include the vehicles themselves. As I see it, Cohen embraces this result (cf. Cohen, 2009, p. 116).<sup>77</sup> But the idea that the neural states involved in perception represent themselves is difficult to square with most of the popular semantic

---

<sup>76</sup> I borrow this way of describing the circularity from Henry Taylor (personal communication).

<sup>77</sup> Here is Cohen openly committing to the visual system's representing itself:

“In a typical perceptual episode, my visual system will begin by representing the lemon as exemplifying the fine-grained property *yellow for S in C* (where ‘S’ is a schematic letter standing in for a relatively detailed specification of my visual system, and ‘C’ is a schematic letter standing in for a relatively detailed specification of the circumstance I am in at the time)” (*ibid.*).

naturalization frameworks, which we got a glimpse of just above.<sup>78</sup> On these frameworks, states' having contents is grounded in the states' causally covarying with distal objects/features. This straightforwardly goes for non-teleosemantic causal theories of content (as we get, e.g., in Fodor, 1987). But it goes for very many teleosemantic theories as well (prominently Dretske, 1988; Neander, 2017; Shea, 2018). Here is the resultant problem. If the vehicles of stygian color contents also constitute stygian color relations, then the vehicles must causally covary with themselves. But nothing causally covaries with itself.<sup>79</sup> So, the vehicles cannot have those contents in a way that can be accommodated by most popular semantic naturalization frameworks.<sup>80</sup>

*Objection: Why not appeal to color pluralism?*

There is this final way the proponent of RR might try to secure the stygian colors' possibility. Color pluralism is the view that it is possible for an object to be distinct colors all over

---

<sup>78</sup> At a certain point, there is nothing odd at all about neural states' representing themselves. After all, we *think* about our neural states. However, the issue here has to do not with cognitively representing our own neural states but with perceptually representing them.

<sup>79</sup> Save perhaps for recherche, and irrelevant, time travel cases—like Effingham and Robson's (2007) self-made, time-travelling brick wall—irrelevant because we should not think content-vehicles causally covary with themselves by way of time travel.

<sup>80</sup> I have left room for stygian color relations' potential accommodation by teleosemantic accounts that do not appeal to causal covariation, accounts like Millikan's (1984) or Papineau's (1993). On these views, a state's content, roughly, is the condition which explains why the state has the effects it does. But it rings false that stygian color relations (the very specific ones mentioned above) might explain why the vehicles of stygian color contents have the effects they do. (Which effects are these supposed to be? Subjects' self-reports?) However, the rest of the case against the naturalizability of visual representation of stygian color relations calls for fuller development elsewhere—if not just for the perhaps surprising fact that the case generalizes beyond the present example to the visual representation of all colors construed relationally.

simultaneously (see Kalderon, 2007, for detailed discussion and defense). Pluralism is compatible with reductionism and primitivism, which is important given the ill-suitedness of relationalism and dispositionalism with RR. Stygian colors would be accommodated by pluralism in the case that it is possible that a single object (or single part of an object) be simultaneously maximally dark all over and hued all over. How exactly would that work on reductionism? Here is Kalderon's suggestion (2007, p. 578). On reductionism, colors are sets/determinables of determinate reflectances.<sup>81</sup> Instantiating a determinate reflectance is a way of being colored—a way of falling under a determinable. For an object to be multiply colored is for its determinate reflectance to belong to distinct determinables, which determinables bear unique similarity-relations to the other color-determinables. By falling under distinct determinables, determinate reflectances “bear different similarity relations to different properties, and so participate in distinct families of properties” (*ibid.*; but see also Kalderon, 2011, sect. 5). Applied to our case, we should say that stygian colors are possible if an object is possibly maximally dark all over and distinctly hued all over, simultaneously.

### *Reply*

This sort of thinking does not look like it will help the proponent of RR. The same considerations which motivated taking stygian colors to be impossible on reductionism in the first place (section 4.1) make intuitive that no determinate reflectance participates in the determinable with which maximal darkness is identical and the determinables with which any hues are identical. The former determinable is the set of determinate reflectances which are basically flat up against the wavelength axes of their associated graphs. The latter are the sets of

---

<sup>81</sup> Note that this is contentious. The reductionist may not want to identify colors with sets of determinate reflectances but identify them instead with, say (recalling fn. 65), whatever as-yet undiscovered feature unifies the members of the set.



determinate reflectances with marked upticks from those axes. We should not expect to find a reflectance that participates in both. These points apply *mutatis mutandis* to primitivism.<sup>82</sup>

### *Conclusion*

The stygian color experiences provide a class of experiences whose members serve as counterexamples to RR. Each stygian experience type lacks a corresponding property which might feature as a constituent in the experience's associated Russellian content. Moreover, treating the experiences' contents as decomposing into constituent elements each of which is possible comes at too steep a price. Accordingly, RR assigns each of the phenomenally distinct stygian color experiences the same content, and the experiences thereby serve as counterexamples to the theory.

The problem posed by the stygian color experiences for RR can be avoided if we posit primitive representational contents, *à la* Burge (2010), and posit these contents' constitution by metaphysically impossible properties. But if we are not antecedently attracted to the existence of metaphysical impossibilities or to primitive representational contents, then the stygian color experiences remain a problem for RR.

---

<sup>82</sup> Thanks to an anonymous reviewer for urging me to consider further what recourse the proponent of RR has to relationalism and pluralism.

## In defense of the response-independence of color

**Abstract:** In this paper I defend the view that colors are response-independent properties from Pautz's argument from structural mismatch (e.g., 2020) and his argument from irregular grounding (e.g., 2016). Each argument depends on the presumption that colors supervene on reflectances. But, I argue, color science is not mature enough at present to warrant that presumption. We need to wait until we discover what (if anything) unifies the metamers before we confidently say what it is color supervenes on. And optimism with respect to discovering what unifies them, I argue, is not ill-advised.

In this paper, I defend the response-independence of color (section 1) against a pair of arguments of Adam Pautz's (section 2). I will argue (section 3) that the proponent of response-independence should say that color science/psychophysics is not mature enough at present to warrant thinking of the putative problem cases as relevant to color's response-independence. The argument depends on the advisability of a certain degree of optimism with respect to the prospect of color scientists' discovering a unifying feature of the metamers (section 4). So I close by motivating why optimism is advised.

### *1 What's at stake*

A property is *response-independent* exactly if its nature can be specified independently of subjects' responses. Having spin-up, for instance, is a response-independent property because whether an electron, for instance, is spin-up does not depend on the responses of subjects (observers in the lab, say). Whereas the property being a paper-weight, for example, is a response-dependent property because whether an object is a paper-weight depends on the intentions or understanding (two species of responses in the relevant sense) of subjects. Returning to the topic of concern, *colors* are response-independent, then, exactly if their natures can be specified independently of subjects' responses.

The question of whether colors specifically are response-independent has been a perennial favorite of metaphysicians at least since the middle Modern period. Galileo famously argued, with other Modern thinkers soon to follow, that colors are properties just of subjects' minds, given then-contemporary physics seemed to exclude them from its image. New trends in metaphysics, however, purportedly in defense of common sense,<sup>83</sup> take the opposing line. Colors do not depend on the mind. They are fully external (response-independent) properties of objects. But here is another reason to care about colors' metaphysical status. Certain prominent *perceptual*

---

<sup>83</sup> See Johnston (1992), Byrne and Hilbert (2003), Churchland (2007), to give just a few examples.

theses hang on the viability of the response-independence of color (hereinafter “RI”). These are externalist perceptual theses, like naïve realism (Campbell, 2009; Martin, 2002; Brewer, 2011; Soteriou, 2013; Logue, manuscript) and extensionalist versions of representationalism (Dretske, 1995; Tye, 2008; Byrne, 2018). Externalist perceptual theses like these ground visual experience of color in response-independent color properties. Cast in terms of phenomenology, on these views what it is like for subjects to enjoy a good case (i.e., a non-flukily veridical case) of visually experiencing something red as such is partly grounded in redness’s response-independent nature. So, these externalist perceptual theses come with them a commitment to RI. Accordingly, an argument against RI is an argument against views like naïve realism and extensionalist representationalism, which are important views in the metaphysics of perception. Very much, then, is at stake.

## 2 Pautz’s argument against RI

In the case that the colors are response-independent properties of objects, the candidate properties to which they might be identical are i) spectral surface reflectances (SSRs), or something thereabouts; or ii) irreducible physical properties of objects that are grounded in SSRs (or something thereabouts). The view which says the colors are identical to SSRs (or something thereabouts) is called *reductionism*. And the latter view is called *primitivism*. These views, today, are the only realist response-independent metaphysical theses of color.<sup>84</sup>

And there is a popular argument against both views of what the colors are. The first part of that argument is intended to show that the colors do not reduce to SSRs. The argument focuses on

---

<sup>84</sup> There are eliminativist versions too. These commit to colors’ being response-independent but uninstantiated. Because these views *prima facie* fail to suit externalist perceptual theses (see the previous essay for the case that eliminativism does not suit extensionalist representationalism), I prescind from them here.

three specific colors and generalizes therefrom. This is the **argument from structural mismatch** (ASM):

1. Whatever blue, purple, and green reduce to, the reduction bases' similarity-structure should reflect the following fact: that blue is more like purple than green.
2. But it is not the case that blue-SSR is more like purple-SSR than green-SSR (see Figure 1).
3. Hence, blue, purple, and green are not SSRs.<sup>85</sup>

Hence, that is, reductionism is false.

The first premise makes subjects' knowledge of the colors' similarity-structure a datum. Our metaphysics of color must accommodate that subjects know about the similarity interrelations of the various colors. We know, for instance, that blue is more like purple than green. We can accept this assumption for the sake of argument. And in philosophy of mind, at least, we usually think of reduction as identity.<sup>86</sup> I will assume this view of reduction in what follows. So, if blue is more like purple than green, then this has to be reflected in the similarity-structure of the things to which these colors are identical.

Reductionism, again, says that the colors reduce to SSRs or something thereabouts. Premise 2, however, disregards the latter disjunct of this formulation (more on which in section 4.1). That is, it is a tacit assumption of the argument that if the colors reduce, then they are identical to SSRs. This is why it is SSRs (those of Figure 1) that are adduced in premise 2.

And if premises 1 and 2 are right, then reductionism is false.

---

<sup>85</sup> See, e.g., Pautz (2020), who uses this name for the argument (pp. 5-6). The argument was introduced to the color metaphysics literature in Hardin (1988)—and has been discussed in many papers since then.

<sup>86</sup> In philosophy of physics it is sometimes said that properties reduce to states of affairs to which they are not identical. Cf. Franklin, 2019, pp. 37-48.

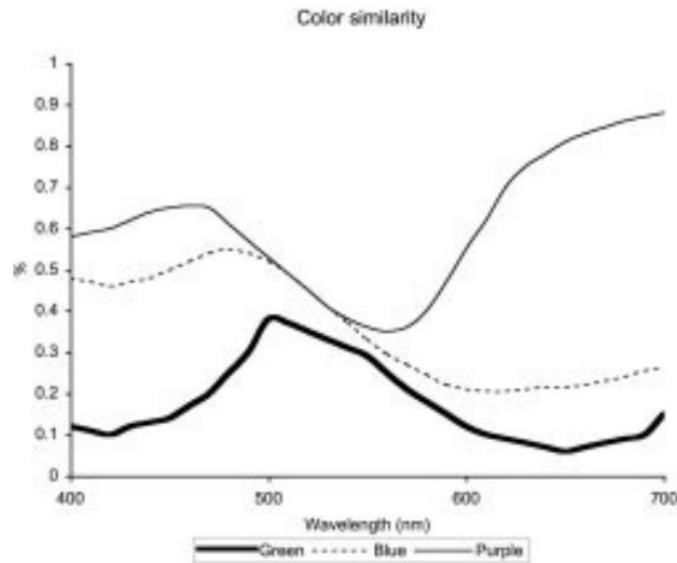


Figure 1. SSRs of typical green, blue, and purple objects. Reprinted with the permission of Cambridge University Press.

But the proponent of RI, we saw above, has recourse to more than just reductionism. Without an argument to the effect that primitivism too is false, the colors stand a chance at being response-independent.

Pautz has argued (2016; forthcoming), that invoking primitivism as a response to the above argument does not work. For invocation of primitivism to avoid the above argument, the would-be primitive would need to commit to the following: i) Blue, purple, and green are grounded in the adduced SSRs (Figure 1). And, ii) even though blue, purple, and green are grounded in those SSRs, the colors are similar and dissimilar in ways which do not mirror the similarity-structure of their grounds (the SSRs).

Pautz says that to commit to (i) and (ii) is to posit “totally unsystematic and arbitrary grounding connections” (2016, p. 28). Whoever accepts (i) and (ii), he goes on, “must hold that this is just a

quirk of reality with no explanation” (*ibid.*). Positing such quirks, he adds, is coherent, “it is just totally implausible” (forthcoming, p. 11).<sup>87</sup>

Putting the above arguments together gives us the following argument against RI, which I will call the **generalized argument from mismatch** (GAM):

1. If primitivism and reductionism fail to respect the ways in which colors are similar and dissimilar, then the views fail as metaphysical accounts of color.
2. If reductionism and primitivism fail as metaphysical accounts of color, then colors are not-response-independent.
3. Colors are similar and dissimilar in ways that reductionism and primitivism fail to respect.
4. Therefore, colors are not response-independent.

### *3 How to respond to the arguments*

I will give my reply first on behalf of reductionism. Then I will apply the case made to primitivism. So, first, here is how the reductionist should respond to the ASM.

#### *3.1 What the reductionist should say*

It is best to say that *color science is at present still not mature enough* for anyone to confidently say which physical properties the colors reduce to. This is primarily due to our having yet to resolve (in a way that is unanimously approved of) the problem of metamerism, which I take it is the greatest obstacle to reducing the colors.<sup>88</sup>

---

<sup>87</sup> I think we should waive the complaint that primitivism is usually framed in terms of supervenience not grounds.

Pautz’s argument rings just as true when the accusation is that of positing “totally unsystematic and arbitrary [supervenience] connections”.

<sup>88</sup> Cf. the introduction of Churchland (2007). For a compendium of the other—more minor—problems, see Hardin’s seminal (1988).

Metamerism is the phenomenon of the colors' having metamers. Two objects' reflectances are metamers exactly when the reflectances are distinct, and yet subjects (in certain specified normal conditions) judge the objects to exhibit the same color. So, if we have two objects  $o_1$  and  $o_2$ , each with respective reflectances  $R_1$  and  $R_2$ , then  $R_1$  and  $R_2$  are metamers of color  $C$  exactly when  $R_1$  and  $R_2$  are not identical and yet subjects, in some set of specified viewing circumstances, judge  $o_1$  and  $o_2$  to be  $C$ . In Figure 2, we see four metamers of yellow.

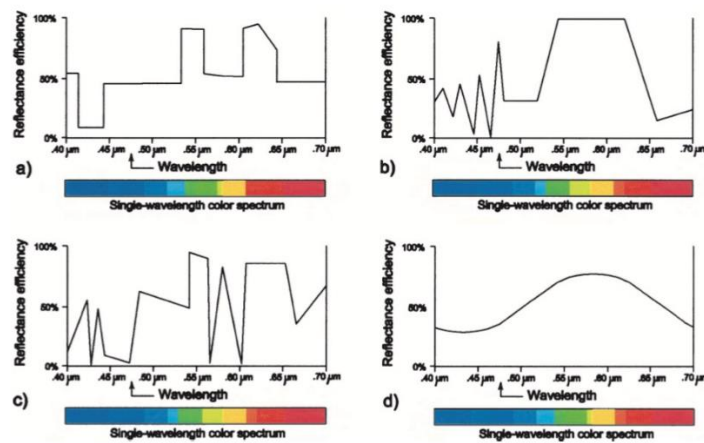


FIGURE 10.1. Four metamers for yellow

Figure 2. Metamers of yellow. Reprinted with the permission of Cambridge University Press.

Metamers are rife, so we cannot treat them as mere oddities or peculiar anomalies. If we could treat them as oddities, then we might say that a given color is identical to whichever SSR is usually present where the color is. Blue, for instance, would be identical to the SSR which is usually present where blue things are. And we could call a color's metamers, given metamers' anomalousness, a sort of *tromp l'oeil*. Metamers for some color would merely be surface properties which subjects are invariably tricked into incorrectly judging to be that color (= to have a certain SSR). Given metamers' ubiquity, this is something we cannot say.

Metamerism would not be an obstacle to reducing the colors if color science had discovered physical properties common to—i.e., which unify—the metamers associated with the colors. If it were that we had discovered which physical property unifies the metamers associated with blue,



for instance, then we would be well on our way to saying blue is identical to that physical property. And so on with the rest of the colors.<sup>89</sup>

Now, there is no unanimous agreement that any attempt at unification has succeeded to date. But Churchland (2007) provides an ingenious such attempt, one that I will argue in section 3.4.2 is yet to face uncontroversially refuting data.

Accordingly, what we need to say is that blueness, purpleness, and greenness do not reduce to the SSRs of Figure 1. If these colors (and the rest) are not identical to reflectance properties yet recognized by color science, then premise 2 of the ASM is irrelevant. The reductionist was not advised to think of SSRs as identical to the colors in the first place. Reductionism, then, is not refuted. And neither, then, is RI taken off the table.

### *3.2 What the primitivist should say*

Pautz, we saw above, advises against identifying the colors with primitive properties to avoid the problem the adduced SSRs present for RI. The thought was that if the colors, though irreducible, are grounded in SSRs, then it looks arbitrary to deny that the similarity-structure of the colors should match that of their grounds. These irregular grounding connections look inexplicable, Pautz says. However, this complaint only works if the colors so conceived are grounded in the adduced SSRs. But the primitivist should not say the colors are grounded in SSRs. There are too

---

<sup>89</sup> Byrne and Hilbert (2003) opt for a different tack, taking the metamers to have no unifying physical features and identifying the colors (hues, really) with disjunctions of metamers. It might seem that if the reductionist went this route, she could deny the relevance of Pautz's adduced SSRs on grounds that the colors are not identical to those (in a manner analogous to the approach I pursue). But there is the following crippling issue with the account, one Pautz himself catches (2003, p. 45). No similarity structure at all captures the interrelations of *disjunctions* of metamers. Because we accepted (above) on behalf of the reductionist that whatever reduces the colors must exhibit the similarity structure colors commonsensically have, the reductionist should not endorse Byrne and Hilbert's proposal as a response to the ASM.

many viable candidates for grounds—the metamers—and no motivation for picking one over the others. If one held that, e.g., yellow is grounded in some one SSR, then what is yellow's relation to the metamers excluded by this choice? If one, for instance, grounded yellow in SSR (d) of the Figure 2, then what is yellow's relation to (a), (b), and (c)? To pick one SSR as grounds to the exclusion of some others would needlessly disqualify many object surfaces as yellow. By that same token, nothing could motivate calling one metamer rather than another the grounds of yellow.

What this means is that if one is sympathetic to primitivism, then one should say the colors are grounded in *whatever it is that the reductionist should say* the colors are identical to.<sup>90</sup> And we have just seen, in section 3.3, that we do not know what that is yet. This means Pautz is not entitled to say that the colors, qua primitive properties, fail to mirror the similarity-structure of their grounds. Therefore, the primitivist does not succumb to Pautz's criticism, and primitivism remains a live response-independent thesis of color.

The upshot of the above is that the fan of RI has recourse to two different theses of color. She can call the colors reducible, or she can call them primitive. It is crucial to note, however, that each option is hostage to empirical fortune. If one wants to endorse the view that the colors reduce, then she must hold out hope for a reduction. This hope would be satisfied, I have said, in the case that the metamers are successfully unified. And in the case that the direct realist wants to call the colors primitive properties: their grounds will be whatever it is that the reductionist is

---

<sup>90</sup> If we succeed in unifying the metamers, this would raise the question of what motivates primitivism. After all, isn't a major motivation for the view precisely the motleyhood of metamers? That might be one motivation, but there are certainly others. For instance, Johnston's idea that color experience reveals colors' natures (Johnston, 1992), together with the thought that unifying features of metamers are not revealed in color experience, is a ready, alternate route to primitivism.

holding out hope to find. Both options, we see, then, are hostage to empirical fortune. And so, both options depend on a bit of optimism.

#### *4 Is optimism advised?*

I want to conclude this section by considering whether this optimism is misplaced—or immodest. Are proponents of RI, like the externalists mentioned above, entitled to hold out hope for a unification of the metamers? I will argue here that they are.

Will Davies (2014) writes that color scientists are unanimous in their expectation that there is not any respect in which the SSRs of Figure 1 will mirror the similarity structure of the colors. Byrne (quoted in Davies, *ibid.*, p. 299) also thinks the hope is misplaced: “[W]e can be completely confident,” he writes, “that any plausible physicalist candidates for the colours do not stand in the required genuine similarity relations” (Byrne, 2003, p. 648).

But the success and value of color science does not hinge on whether the colors are response-independent features of reality. Most color scientists do not believe that the colors exist externally at all (Pautz, 2020, p. 22). So, it is not especially suggestive that color scientists tend to be pessimistic about a consequence (structural match) of an idea (response-independence) that the success and value of their field does not depend on.

What I want to argue is that optimism is advised because of the potential success of Churchland’s (2007) proposed unification of the metamers. In that paper, Churchland develops a method for extracting mathematical features of SSRs where the extracted features i) are similar in just the ways the colors are commonsensically taken to be and which ii) are hypothesized to unify the classes of metamers. Finding features of SSRs that satisfy both (i) and (ii) constitutes an irresistible motivation for reducing the colors to those features. (Here, an important qualification needs making, one we can leave for later.) I turn to Churchland’s framework now.

#### *4.1 Churchland’s canonical approximation ellipses*

Churchland's hypothesis is that it is the *canonical approximations* of metamers which unify them. To find a metamer's (SSR's) canonical approximation (CA), we must first take a graphic representation of a particular SSR. As we have seen (Figure 1, Figure 2), these are rectangular graphs the x-axes of which denote electromagnetic radiation wavelength, starting with 400nm and ending at 700nm, and the y-axes of which denote how strongly radiation of a certain wavelength is reflected. One is to roll these rectangular graphs into cylinders so that the beginning of the x-axis (400nm) meets its end (700nm). (If you have a paper graphic-representation of a reflectance, you can roll it into such a cylinder by cutting the graph out of the page.) The CA of the reflectance is a planar slice, an ellipse, through the resultant cylinder which meets the following conditions:

- “1. the altitude of the ellipse must be such that the total area  $A$  above the [slice], but below the several upper reaches of the target reflectance profile, is *equal* to the total area  $B$  beneath the [slice], but above the several lower reaches of the target reflectance profile. (This condition guarantees that the *total area* under the target reflectance profile equals the total area under the [slice].)
2. The angle by which the [slice] is tilted away from the horizontal plane, and the rotational or compass-heading positions of its upper extreme, must be such as to *minimize* the magnitude of the two areas  $A$  and  $B$ . (This condition guarantees that the [slice] *follows* the gross shape of the target reflectance profile, at least to the degree possible)” (208, italics original).

Each class of metamers has one CA ellipse, Churchland says. And, so, all of the metamers of yellow (four of which we see in Figure 2), for instance, will be united by a single CA ellipse. Each yellow-metamer has just one CA ellipse, and all yellow-metamers have that same CA ellipse.

This, if correct, provides tremendous support for reductionism. I have said that the reductionist could not reduce the colors until we have unified the metamers. CA ellipses look, then, like a viable candidate for the reduction base of the colors. But we have said also that we want to show

that the similarity-structure of the colors mirrors the similarity-structure of the reduction base. If CA ellipses differ in ways that do not mirror the similarity-structure of the colors, then an argument could be made which exactly parallels the argument from structural mismatch.

Specifically, if the CA ellipses corresponding to blue, purple, and green do not mirror those three colors' similarity-structure, premise 2 of that parallel argument would say: "But it is not the case that the blue CA ellipse resembles the purple CA ellipse more than the green CA ellipse."

Churchland proposes that these are the dimensions of the relevant reduction base—structure: the CA ellipses' altitudes (as determined by their centers); their tilt angles; and their rotational positions (pp. 210-11). And the structure comprised of these dimensions (p. 211) indeed mirrors the similarity-structure the colors—all of them—commonsensically have.<sup>91</sup>

That the metamers are unified by CA ellipses speaks in favor of reducing the colors to them.

That there are also objective dimensions of difference along which the CA ellipses can be arranged which generates the above structure-matching makes that reduction irresistible. And, if we think of the colors as reducing to Churchland's CA ellipses, then we can dismiss as irrelevant premise 2 of the ASM: it does not matter in what ways the adduced SSRs differ, for those are not the colors.

Before we move on, a qualification of the proposed reduction is in order. Churchland himself thinks it is a bad idea to identify the colors with his CA ellipses (2007, p. 221). He says, "Simply identifying the familiar range of colors with the evident range of CA ellipses is a very poor option, since only a negligible proportion of material objects have a reflectance profile that is actually identical with the Platonic perfection of a CA ellipse" (*ibid.*). But this remark neglects the following obvious alternative. The above discussion speaks in favor of reducing the colors not

---

<sup>91</sup> Where this mirroring is not strict isomorphism but something more like homomorphism, or perhaps an even more relaxed notion of structural resemblance. See Churchland, *ibid.*, sect. VI.

CA ellipses but to *objects' having* such-and-such CA ellipses. What we should be saying is: for an object to be, for instance, yellow is for it to have the yellow CA ellipse. All yellow things, Churchland hypothesizes, do have the yellow CA ellipse, even if their exact reflectance is not itself the CA ellipse. For an object to have a certain CA ellipse is for its (exact) reflectance to be treatable in the way discussed above. So the reduction is not unavailable for the reason he notes.

#### 4.2 Counterexamples to Churchland's proposal

Kuehni and Hardin (2010) attempt to show that Churchland's framework gets the wrong results in a handful of cases. There are metamers of certain colors, they show, which Churchland's framework mislabels.

Kuehni and Hardin provide, first, three metamers of gray which Churchland's framework fails to unify. That is, the gray-metamers have different CAs. Churchland's proposal, then, entails that they are different colors. Specifically, his framework entails that one should appear a bluish-green and another a yellowish-red (2010, sect. 5.1). And then they provide three metamers of yellow the CAs of which are dramatically unequal (sect. 5.2, Table 2). If these examples are appropriate, then it will have been shown that Churchland's framework does not generalize beyond the handful of cases on which he tested his proposal (Churchland, 2007, sect. V).

But the following point may disqualify their counterexamples. Kuehni and Hardin cull their metamers from the *CIE space of colors* (Kuehni and Hardin, 2010, p. 89). The CIE space of colors is a color space that models the subjective responses of standard perceivers to different combinations of light.

But in discussing work of L. D. Griffin's (2001) to establish a mirroring of the similarity-structure of the CIE space and the commonsense similarity-structure of the colors, Churchland is emphatic that the CIE space is not a good candidate reduction base of the colors. He writes,

“The CIE space is a space for representing and analyzing *illuminants*, not reflectance profiles . . . . It is a perfectly good and useful space, but it does not address the reality of the objective colors of the vast majority of objects in our terrestrial environment, which are almost exclusively *reflectance* colors, not self-luminous colors. Moreover, it fails to represent the all-important dimension of objective lightness and darkness captured by the space of possible CA-ellipses. . . . The CIE space has no room for black, for example, or for any of the darkish colors in the neighborhood of black.” (2007, p. 217 italics original).

And, so, perhaps it is inappropriate to exploit cases from the CIE space if one hopes to counterexemplify Churchland’s proposal. Churchland explicitly and emphatically challenges the appropriateness of taking the colors to reduce to regions of that space. His proposal is geared specifically to reflectances, aspects of these being the best candidate reduction base for the colors.<sup>92</sup>

#### *4.3 Cause for optimism*

Now, that we have not yet seen data which clearly disconfirms Churchland’s proposal does not mean we must accept it. Churchland tests his proposal against just three groupings of metamers (two in sect. V; and I am assuming he also tested his hypothesis against the metamers of Figure 10.1 p. 201—our Figure 2 above). So, there might be (less contentious) counterexamples in the

---

<sup>92</sup> The above reply, then, is hostage to categorizing emittance colors as fundamentally distinct in kind from reflectance colors (Churchland, 2007, sect. VIII). Churchland thinks this is an independently motivated distinction to make. But I imagine many philosophers will be antecedently attracted to Byrne and Hilbert’s (*op. cit.*) “productance” view of colors will be hesitant to embrace this distinction. But the matter must be left for another time.

offing. And in any case, there are more challenges to Churchland's proposal than just Kuehni and Hardin's.<sup>93</sup>

However, though this may recommend against our racing to accept his proposal, what we have yet to be given is cause for pessimism regarding the prospects of a unification of the metamers. If anything, the potential success of Churchland's proposal gives us cause to be sanguine indeed. Accordingly, the fan of RI, of reductionist or primitivist stripe, is not advised against committing to a reply to the GAM which is hostage to empirical fortune.

### **Upshot**

The upshot of the above is that we should safely be able to avoid the thrust of the GAM. We do not yet know what the colors are identical to or grounded in. Moreover, we are entitled to optimism regarding the prospects of finding a viable candidate for such a reduction base/grounds. After all, it may be that we have one such candidate now. Accordingly, we have yet to be given a problem case involving those bases'/grounds' differing in ways which fail to mirror the similarity-structure colors are commonsensically taken to have. And for that, reductionism and primitivism—and, so, RI—remain live options. And this means Pautz has yet to give us a reason to abandon externalist theses of perceptual experience, iterations of which depend on RI.

---

<sup>93</sup> There are, for instance: Isaac (2009) and Wright (2009); and Kuehni and Hardin challenge other aspects of Churchland's proposal in their (*op. cit.*). I take these further challenges to be less pressing than the ones discussed above.



## Bibliography

Alston, W. (1999) 'Back to the Theory of Appearing', *Philosophical Perspectives*, 13: 181–203.

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.

Beck, Ori (2019). Rethinking naive realism. *Philosophical Studies* 176 (3):607-633.

Block, Ned (1995). On a confusion about a function of consciousness. *Brain and Behavioral Sciences* 18 (2):227–247.

Block, Ned (1996). Mental paint and mental latex. *Philosophical Issues* 7:19-49.

Block, Ned (2010). Attention and mental paint. *Philosophical Issues* 20 (1):23-63.

Block, Ned (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research*. 89: 560–572. 2016.

Block, Ned (2016). The Anna Karenina Principle and Skepticism about Unconscious Perception. *Philosophy and Phenomenological Research* 93 (2):452-459.

Block, Ned and Phillips, I. (2016). Block/Phillips Debate on Unconscious Perception. In B. Nanay (ed.) *Current Controversies in Philosophy of Perception* New York: Routledge.

Brewer, Bill (2007). Perception and its objects. *Philosophical Studies* 132 (1):87-97.

Brewer, Bill (2011). *Perception and Its Objects*. Oxford University Press.

Brewer, Bill ; de Bruijn, David ; Hill, Chris ; Pautz, Adam ; Rosenhagen, Raja ; Vuletić, Miloš & Wu, Wayne (2018). Discussion of Bill Brewer's "Perceptual Experience and Empirical Reason". *Analytic Philosophy* 59 (1):19-32.

Brouwer, G. and D. Heeger. 2013. "Categorical Clustering of the Neural Representation of Color." *Journal of Neuroscience* 33: 15454–65.

Burge, Tyler. (2010). *Origins of Objectivity*. Oxford University Press.

Byrne, Alex & Hilbert, David R. (2003). Color realism and color science. *Behavioral and Brain Sciences* 26 (1):3-21.

- Byrne, Alex & Hilbert, David R. (2006). Color primitivism. In Ralph Schumacher (ed.), *Erkenntnis*. Kluwer Academic Publishers. pp. 73 - 105.
- Byrne, Alex & Logue, Heather (2008). Either/or. In Adrian Haddock & Fiona Macpherson (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press. pp. 314-19.
- Byrne, Alex (2003). Color and similarity. *Philosophy and Phenomenological Research* 66 (3):641-65.
- Byrne, Alex (2006). Color and the Mind-Body Problem. *Dialectica* 60 (2):223-44.
- Byrne, Alex (2018). *Transparency and Self-Knowledge*. Oxford University Press.
- Campbell, John (1993). A simple view of colour. In John J. Haldane & C. Wright (eds.), *Reality: Representation and Projection*. Oxford University Press. pp. 257-268.
- Campbell, John (2002). *Reference and Consciousness*. Oxford University Press.
- Campbell, John (2009). Consciousness and Reference. In Brian McLaughlin, Ansgar Beckermann & Sven Walter (eds.), *The Oxford Handbook of Philosophy of Mind*. Oxford University Press.
- Carruthers, P. (2000). *Phenomenal consciousness: A naturalistic theory*. Cambridge: Cambridge University Press.
- Chalmers, David J. (2010). *The Character of Consciousness*. Oxford University Press
- Chalmers, David J. (2018). The Meta-Problem of Consciousness. *Journal of Consciousness Studies* 25 (9-10):6-61.
- Churchland, Paul (2005). Chimerical colors: Some phenomenological predictions from cognitive neuroscience. *Philosophical Psychology* 18 (5):527-560.
- Churchland, Paul (2007). On the reality (and diversity) of objective colors: How color-qualia space is a map of reflectance-profile space. *Philosophy of Science* 74 (2):119-149.
- Churchland, Paul M. (2012). *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. MIT Press.

- Churchland, Paul M. (2012). *Plato's Camera: How the Physical Brain Captures a Landscape of Abstract Universals*. MIT Press.
- Cohen, Jonathan (2004). Color properties and color ascriptions: A relationalist manifesto. *Philosophical Review* 113 (4):451-506.
- Cohen, Jonathan (2009). *The Red and the Real: An Essay on Color Ontology*. Oxford University Press UK.
- Conee, Earl (1994). Phenomenal knowledge. *Australasian Journal of Philosophy* 72 (2):136-150.
- Crane, Tim and Craig French, "The Problem of Perception", *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2021/entries/perception-problem/>.
- Davidson, Donald (1963). Actions, Reasons, and Causes. *Journal of Philosophy* 60 (23):685.
- Davies, Will (2014). The inscrutability of colour similarity. *Philosophical Studies* 171 (2):289-311.
- Dennett, Daniel C. (1988). Quining qualia. In Anthony J. Marcel & E. Bisiach (eds.), *Consciousness in Modern Science*. Oxford University Press.
- Dennett, Daniel C. (2005). *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*. MIT Press.
- Dennett, Daniel C. (2016). Illusionism as the Obvious Default Theory of Consciousness. *Journal of Consciousness Studies* 23 (11-12):65-72.
- Dretske, Fred (1995). *Naturalizing the Mind*. MIT Press.
- Dretske, Fred I. (1988). *Explaining Behavior: Reasons in a World of Causes*. MIT Press.
- Fish, William C. (2008). Disjunctivism, indistinguishability, and the nature of hallucination. In Adrian Haddock & Fiona Macpherson (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press. pp. 144--167.
- Fish, William C. (2009). *Perception, Hallucination, and Illusion*. Oxford University Press.

- Fish, William C. (2021). Perceptual paradigms. *Purpose and Procedure in Philosophy of Perception*, eds. Heather Logue and Louise Richardson. Oxford University Press.
- Fodor, J. A., 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: MIT/Bradford.
- Fraklin, Alexander (2019). Why isn't there only physics? PhD thesis, King's College London, London.
- Frankish, Keith (2012). Quining diet qualia. *Consciousness and Cognition* 21 (2):667-676.
- Frankish, Keith. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies* 23 (11-12):11-39.
- French, Craig (2014). Naive Realist Perspectives on Seeing Blurrily. *Ratio* 27 (4):393-413.
- French, Craig (2020). Naïve realism, representationalism, and the rationalizing role of visual perception. *Philosophical Issues* 30, Issue 1.
- Gow, Laura (2016). The Limitations of Perceptual Transparency. *Philosophical Quarterly* 66 (265):723-744.
- Green, E.J. (2020). Representing Shape in Sight and Touch. *Mind & Language*.
- Green, E.J. (2021). The Puzzle of Cross-Modal Shape Experience. *Noûs*.
- Griffin, L. D. (2001). "Similarity of Psychological and Physical Color Space Shown by Symmetry Analysis" *Color: Research and Application* 26, no. 2: 151–7.
- Grimm, Stephen R. (2016). How Understanding People Differs from Understanding the Natural World. *Philosophical Issues* 26 (1):209-225.
- Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Hackett.
- Harman, G. (1990) 'The Intrinsic Quality of Experience', in J. Tomberlin (ed.) *Philosophical Perspectives* 4/ Action Theory and Philosophy of Mind, 31–52. Atascadero, CA: Ridgeview.

- Holman, E. 2002. Color Eliminativism and Color Experience. *Pacific Philosophical Quarterly* 83: 8–56.
- Huemer, Michael, "Sense-Data", *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/sense-data/>.
- Isaac, Alistair (2009). Prospects for naturalizing color. *Philosophy of Science* 76 (5):902-914.
- Jackson, Frank (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32 (April):127-136.
- Jakab, Z. 2003. Phenomenal Projection. *Psyche* 9 (4).
- Johnston, Mark (1992). How to speak of the colors. *Philosophical Studies* 68 (3):221-263.
- Kalderon, Mark Eli (2007). Color pluralism. *Philosophical Review* 116 (4):563-601.
- Kalderon, Mark Eli (2011). Color Illusion. *Noûs* 45 (4):751-775.
- Kalderon, Mark Eli (2011). The Multiply Qualitative. *Mind* 120 (478):239-262.
- Kammerer, François (2016). The hardest aspect of the illusion problem - and how to solve it. *Journal of Consciousness Studies* 23 (11-12):124-139.
- Kammerer, François (2018). Can you believe it? Illusionism and the illusion meta-problem. *Philosophical Psychology* 31 (1):44-67.
- Kaplan, Mark (1981). Rational acceptance. *Philosophical Studies* 40 (2):129 - 145.
- Kriegel, Uriah (2013) The phenomenal intentionality research program. *Phenomenal Intentionality*. ed. Uriah Kriegel. Oxford University Press.
- Kripke, Saul (1980). *Naming and Necessity: Lectures Given to the Princeton University Philosophy Colloquium*. Cambridge, MA, USA: Harvard University Press.
- Kuehni, Rolf G. & Hardin, C. L. (2010). Churchland's metamers. *British Journal for the Philosophy of Science* 61 (1):81-92.
- Langsam, H. (1997) 'The Theory of Appearing Defended', *Philosophical Studies*, 87: 33–59.

- Levin, Janet (2000). Dispositional theories of color and the claims of common sense. *Philosophical Studies* 100 (2):151-174.
- Lewis, D., 1988, "What Experience Teaches", *Proceedings of the Russellian Society*, 13: 29–57; reprinted in W. G. Lycan, 1990b, 499–519, and in P. Ludlow, et al, 2004, 77–103.
- Livitz, G. et al. (2011). Perceiving Opponent Hues in Color Induction Displays.' *Seeing and Perceiving* 24, 1-17.
- Loar, Brian (1990/1997). Phenomenal states. *Philosophical Perspectives* 4: 81-108. Revised in N. Block, O. Flanagan, and G. Güzeldere, eds., *The Nature of Consciousness*. Cambridge, Mass.: MIT Press).
- Loar, Brian (2002). Transparent experience and the availability of qualia. In Quentin Smith & Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Loar, Brian (2004). Phenomenal States (Revised Version). In Yujin Nagasawa, Peter Ludlow & Daniel Stoljar (eds.), *Philosophical Perspectives*. MIT Press. pp. 219.
- Logue, Heather (2013). Good News for the Disjunctivist about (one of) the Bad Cases. *Philosophy and Phenomenological Research* 86 (1):105-133.
- Logue, Heather (forthcoming). Manuscript.
- Lycan, William (1996). *Consciousness and Experience*. MIT Press.
- Lycan, William (2008). Phenomenal intentionalities. *American Philosophical Quarterly* 45 (3):233 - 252.
- Lycan, William (2019). Block and the representation theory of sensory qualities. *Blockheads! Essays on Ned Block's Philosophy of Mind and Consciousness*. Eds. Pautz, Adam & Stoljar, Daniel. New York: MIT Press.
- Lycan, William G. (1996). *Consciousness and Experience*. MIT Press.
- Machery, Edouard (2009). *Doing Away with Concepts*. Oxford University Press.

- Macpherson, Fiona (2003). Novel colours and the content of experience. *Pacific Philosophical Quarterly* 84 (1):43-66.
- Martin, M. G. F. (2002). "The transparency of experience." *Mind and Language*, 17: 376-425.
- Martin, M. G. F. (2004). 13 The Limits of Self-Awareness. In Heather Logue & Alex Byrne (eds.), *Disjunctivism: Contemporary Readings*. MIT Press. pp. 271.
- Martin, M. G. F. (2006). On being alienated. In Tamar S. Gendler & John Hawthorne (eds.), *Perceptual Experience*. Oxford University Press.
- Martin, M. G. F. (2010). What's in a look? In Bence Nanay (ed.), *Perceiving the World*. Oxford University Press. pp. 160-225.
- Masrour, Farid (2020). On the Possibility of Hallucinations. *Mind* 129 (515):737-768.
- Maund, J. B. (1995). *Colours: Their Nature and Representation*. New York: Cambridge University Press.
- McDowell, J. (1985). "Values and Secondary Qualities." In Ted Honderich (ed.) *Morality and Objectivity: A Tribute to J. L. Mackie*, London: Routledge and Kegan Paul.
- McGinn, Colin (1996). "Another Look at Color", *The Journal of Philosophy*, 93(11): 537–555.  
doi:10.2307/2941048
- Millikan, Ruth Garrett (1984). *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press.
- Nagel, Thomas (1974). What is it like to be a bat? *Philosophical Review* 83 (October):435-50.
- Neander, Karen (2017). *A Mark of the Mental: A Defence of Informational Teleosemantics*. Cambridge, USA: MIT Press.
- Nemirow, L., 1980, "Review of Thomas Nagel, *Mortal Questions*," *Philosophical Review*, 89: 473–477.

- Opie, Jonathan & O'Brien, Gerard (2004). Notes toward a structuralist theory of mental representation. In Hugh Clapin, Phillip Staines & Peter Slezak (eds.), *Representation in Mind: New Approaches to Mental Representation*. Elsevier. pp. 1--20.
- O'Shaughnessy, Brian, 2003. "Sense Data," pp. 169–88 in *John Searle*, ed. Barry Smith. Cambridge: Cambridge University Press.
- Papineau, David (1993). *Philosophical Naturalism*. Wiley-Blackwell.
- Papineau, David (2021). *The Metaphysics of Sensory Experience*. Oxford: Oxford University Press.
- Pautz, Adam (2003). Have Byrne & Hilbert answered Hardin's challenge? *Behavioral and Brain Sciences* 26 (1):44-45.
- Pautz, Adam (2016). Experiences are Representations: An Empirical Argument (forthcoming Routledge). In Nanay (ed.), *Current Controversies in the Philosophy of Perception*. Routledge.
- Pautz, Adam (2020). How Does Colour Experience Represent the World? In *Routledge Handbook of the Philosophy of Colour*. Routledge.
- Pautz, Adam (forthcoming). Naive Realism and the Science of Consciousness. *Analytic Philosophy*.
- Pereboom, D. (2016). Illusionism and Anti-Functionalism about Phenomenal Consciousness. *Journal of Consciousness Studies* 23 (11-12):172-185.
- Perry, John. (2001). *Knowledge, Possibility, and Consciousness*. Cambridge, Mass.: MIT Press.
- Phillips, Ian (2016). Consciousness and Criterion: On Block's Case for Unconscious Seeing. *Philosophy and Phenomenological Research* 93 (2):419-451.
- Phillips, Ian (2018). Unconscious Perception Reconsidered. *Analytic Philosophy*.
- Place, Ullin T. (1956). Is consciousness a brain process. *British Journal of Psychology* 47 (1):44-50.
- Putnam, H. (1999). *The Threefold Chord*. New York: Columbia University Press.
- Robinson, Howard, 1994. *Perception*, London: Routledge.
- Ryder, Dan (2004). SINBaD neurosemantics: A theory of mental representation. *Mind and Language* 19 (2):211-240.



- Shea, Nicholas (2018). *Representation in Cognitive Science*. Oxford University Press.
- Shoemaker, S. (1994). Self-knowledge and 'Inner-sense'. *Philosophy and Phenomenological Research* 54, pp. 249–314
- Siegel, Susanna (2004). Indiscriminability and the phenomenal. *Philosophical Studies* 120 (1-3):91-112.
- Siegel, Susanna (2008). The Epistemic Conception of Hallucination. In Adrian Haddock & Fiona Macpherson (eds.), *Disjunctivism: Perception, Action and Knowledge*. Oxford University Press.
- Smith, A. D. (2008). Disjunctivism and discriminability. In Adrian Haddock & Fiona Macpherson (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press.
- Snowdon, Paul (2008). Hinton and the origins of disjunctivism. In Adrian Haddock & Fiona Macpherson (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press.
- Soteriou, Matthew (2013). *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford University Press.
- Speaks, Jeff (2015). *The Phenomenal and the Representational*. Oxford University Press UK.
- Stoljar, Daniel (2010). *Physicalism*. Routledge.
- Stoljar, Daniel & Nagasawa, Yujin (2004). Introduction to *There's Something About Mary*. In Peter Ludlow, Daniel Stoljar & Yujin Nagasawa (eds.), *There's Something About Mary*. Cambridge, MA.
- Sturgeon, Scott (2000). *Matters of Mind: Consciousness, Reason and Nature*. London: Routledge.
- Sturgeon, Scott (2006). Reflective disjunctivism. *Proceedings of the Aristotelian Society*, 106: 185-216.
- Sturgeon, Scott (2008). Disjunctivism about visual experience. In Adrian Haddock & Fiona Macpherson (eds.), *Disjunctivism: Perception, Action, Knowledge*. Oxford University Press. pp. 112--143.
- Taylor, Henry (2020). Fuzziness in the Mind: Can Perception be Unconscious? *Philosophy and Phenomenological Research* 101 (2):383-398.

- Thau, Michael (2002). *Consciousness and Cognition: Unified Account*. Oxford University Press.
- Travis, Charles (2004). The silence of the senses. *Mind*, 113(449):57–94, January 2004.
- Tye, M. (2002). Visual qualia and visual content revisited. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (pp. 447-456). New York: Oxford University Press.
- Tye, Michael (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. MIT Press.
- Tye, Michael (2000). *Consciousness, Color, and Content*. MIT Press.
- Tye, Michael (2003). Blurry images, double vision, and other oddities: New problems for representationalism. In Quentin Smith & Aleksandar Jokic (eds.), *Consciousness: New Philosophical Perspectives*. Oxford University Press.
- Tye, Michael (2009). *Consciousness Revisited: Materialism Without Phenomenal Concepts*. MIT Press.
- Tye, Michael (2014). 'Transparency, Qualia Realism and Representationalism'. *Philosophical Studies*, 170: 39–57.
- Tye, Michael, "Qualia", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/sum2018/entries/qualia/>>.
- Williamson, Timothy (1990). *Identity and Discrimination*. Wiley-Blackwell.
- Williamson, Timothy (2013). How Deep is the Distinction between A Priori and A Posteriori Knowledge? In Albert Casullo & Joshua C. Thurow (eds.), *The A Priori in Philosophy*. Oxford University Press. pp. 291-312.
- Wright, Wayne (2003). Projectivist Representationalism and Color. *Philosophical Psychology* 16: 515–33.
- Wright, Wayne (2009). Discussion: The Physical Unnaturalness of Churchland's Ellipses. *Philosophy of Science* 76 (3):391-403.