HALLUCINATION AND NAIVE REALISM

By

KARSON KOVAKOVICH

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Ernest Sosa

And approved by

____________________________________
____________________________________
____________________________________
____________________________________

New Brunswick, New Jersey

January 2022
ABSTRACT OF THE DISSERTATION

Hallucination and Naive Realism

By KARSON KOVAKOVICH

Dissertation Director:

Ernest Sosa

A perennial challenge for naive realism is the argument from hallucination which relies on some version of the claim that veridical perception and hallucination share some deep similarity. Naive realists account for the character of veridical experience in terms of certain special relations to parts of the external world. But some hallucinations seem to share the same character despite those relations being absent. In recent literature, Michael Martin has sharpened this worry about hallucination by presenting the so-called ‘screening off problem’. The idea is that the causal factors that explain the character of experience in the case of hallucination are also present in the case of veridical perception, and that such factors should be adequate to explaining the character of experience in both cases (since things look the same to a subject as between a case of veridical perception and its hallucinatory counterpart). Since, according to the argument, such factors do not involve the naive realist’s special relations to the external world, such relations are explanatorily screened off.

In this dissertation, I carefully examine the range of naive realist theories on the market and attempt to extract some core theses and thereby unify presentations that differ in terminology. As important ground clearing, I articulate along the way the various relevant kinds of perceptual experience and the crucial ideology of phenomenal
character. Next, I examine what is often regarded as the most powerful case for naive realism, namely its intuitive appeal and argue that the supposed intuitive advantages of naive realism over its main competitor, intentionalism, are overstated.

In evaluating the screening off problem, I argue, its standard presentation is quite problematic. I go on to present a new variant of the screening off problem, one based on narrowness rather than causation, that is crisper and more challenging than the original version. I present a menu of escape routes for the naive realist. The good news for the naive realist is that there are quite a few possible escape routes. The bad news is that they are all unappetizing.

Finally, I turn to William Fish’s naive realism, which constitutes the most prominent recent defense of that approach. Two novel arguments are presented against Fish’s account of hallucination. One of those arguments, which receives extended attention, turns on the HP principle: that hallucinability entails possibility. First, I show that Fish’s view is committed to the HP principle. Next, I work through several different potential counterexamples and use them as case studies to test the HP principle. I ultimately argue against the HP principle and thereby demonstrate the failure of Fish’s positive account of hallucination. I also argue that Fish imposes indefensible connections between the character of experience and rational agents.
Acknowledgements

My greatest philosophical debt is to John Hawthorne. His intellectual mentorship over my circuitous academic path has been invaluable. If there is any helpful insight provided in this dissertation, it is largely due to John’s guidance. I am grateful for the encouragement and support from my Rutgers committee members: Ted Sider, Ernie Sosa, and Dean Zimmerman. Special thanks to Ted and the Rutgers philosophy department for allowing me to finish the PhD after a long hiatus.

I’m indebted to Eric Funkhouser and Adam Sennet, who each provided comments on earlier stages of this dissertation. Over the years, I’ve benefited from innumerable discussions on this topic with a great many people, with varying philosophical, scientific, and religious backgrounds. If you are one of those people, I thank you.

This dissertation does not happen without the love and support of my family and friends. I thank my beloved spouse, Lisa, for her unwavering emotional and spiritual support and enabling me to write through the pandemic. My gratitude is eternal. I dedicate this dissertation to my grandmother, Marilyn, who taught me about what matters most and always encouraged my philosophical endeavors.
# Table of Contents

Abstract ii  
Acknowledgements iv  

1 Naive Realism 1  
  1.1 Introduction 2  
  1.2 The Categories of Perceptual Experience 6  
  1.3 The Phenomenal Character of Experience 18  
  1.4 Naive Realism 30  
  1.5 Veridical Phenomenology 36  
  1.6 Concluding Remarks 54  

2 Intuitive Appeal 56  
  2.1 Preliminaries 56  
  2.2 Presentationism and Directness 61  
  2.3 Vulgar Quietism on Fundamentality and Externality 66  
  2.4 Pre-Theoretical Reflection and Phenomenology 66  
  2.5 Concluding Remarks 72  

3 Hallucination and the Screening Off Problem 73  
  3.1 Preliminaries 74  
  3.2 A Challenge Concerning the Entities of Perceptual Awareness 79  
  3.3 The Screening Off Problem 82  
  3.4 Replies to the Screening Off Problem 92  
  3.5 Concluding Remarks 112  

4 Phenomenal Disjunctivism 113  
  4.1 Disjunctivism about Phenomenal Character 114  
  4.2 Rationality and Hallucination 125  
  4.3 Hallucinability and Possibility 127  
  4.4 Reasons Against Hallucinability Entailing Possibility 132  
  4.5 Concluding Remarks 154  

References 155
Chapter 1
Naive Realism

Naive realism is a theory about veridical perception. Naive realists provide an account of the nature of veridical perceptual experience, and the phenomenal character of such experience, in terms of fundamental perceptual relations obtaining between perceivers and ordinary mind-independent objects (and their properties). A minimal condition on a general philosophical theory of perceptual experience is that it provides an account of each of the categories of perceptual experience, not just veridical perception. In this dissertation, I will be primarily concerned with theories of hallucination compatible with naive realism about veridical perception.

In this chapter, I begin by providing a brief introduction to the topics covered in this dissertation and the plan forward (1.1). Next, I cover preliminary matters that include developing theory-neutral definitions of the categories of perceptual experience built out of theoretically-lightweight notions of perceptual success and failure (1.2) and also developing a framework for understanding the phenomenology of experience through an axiomatic account of the key concepts (1.3). The primary objective in this chapter is to present the core theses of naive realism (1.4), which is followed by an extended discussion of the naive realist account of the phenomenal character of veridical perceptual experience (1.5). I end the chapter with a summary highlighting the key results (1.6).
1.1 Introduction

Naive Realism has been on the rise since the turn of the 21st century. While often regarded as the point of departure—encapsulating our ordinary view—in the philosophy of perception, naive realism struggles to adequately account for perceptual error. This is the so-called “problem of perception” and its accompanying arguments from illusion and hallucination. Following most of the literature on naive realism, my primary focus will be on hallucination.

Disjunctivism, as a theory of perceptual experience, is a relatively recent and novel reply to the argument from hallucination. Disjunctivism first appears in the literature in Hinton (1967a, 1967b, 1973). The earliest subsequent discussions are found in McDowell (1983, 1986) and Snowdon (1980, 1990). This dissertation is solely concerned with disjunctivism about the metaphysical nature of perceptual experience and will not be addressing epistemological disjunctivism about the nature of rational support or evidence. Henceforth, any unqualified use of ‘disjunctivism’ refers to metaphysical disjunctivism about perceptual experience.

Disjunctivism rejects the premise that a veridical perception and a subjectively indistinguishable hallucination have the same nature. This opens up the possibility of preserving naive realism about veridical perception while providing another account for hallucination.\footnote{We could define disjunctivism in general terms such that it does not entail naive realism. But since disjunctivism is being developed as a way of defending naive realism, I will use ‘disjunctivism’ to label any view that endorses naive realism about veridical perception and a different view about the nature of hallucination. Illusion might be treated separately as a third view or one more like one of the others. It is worth noting that a disjunctivist might put forward many different views in order to account for different varieties of hallucination (and illusion).} By contrast, if they did have the same nature, then it is hard to see how the nature of veridical experience could be given in terms of standing in relations to ordinary
objects in the environment. Thus, a disjunctivist account of perceptual experience rejects the constraint that veridical perception and hallucination are of a common kind with a shared ontological structure.

It is widely held that any account of hallucination, proffered by a disjunctivist, is highly constrained by an influential argument by Martin (2004, 2006) called the screening-off problem. The crux of the problem is this: whatever positive account is provided of the phenomenal character of hallucinatory experience, it is possible the same phenomenal character of a hallucinatory experience is present in a subjectively indistinguishable case of veridical perception and so the positive account of hallucination explanatorily “screens off” the naive realist account of veridical perception. The argument proceeds by showing that there are always phenomenal properties of a hallucinatory experience present in cases of veridical perception. Thus, the threat is that the naive realist account of veridical perception, specifically the account of the phenomenal character of veridical perception, may become explanatorily impotent if a positive account of hallucination is given. Thus, the screening-off problem presents an obstacle to the disjunctivist in providing a positive account of hallucination.

Martin’s disjunctivist response to the screening off problem, evidential disjunctivism, is to only provide a negative account of hallucination. His negative account entails the naive realist account of veridical perception (Martin 2004, 2006, 2013). A second response to the screening-off problem, which I will also categorize as a negative account, is to deny that total hallucinations have phenomenal character. This radical strategy, phenomenal disjunctivism, is pursued by Fish (2008, 2009, 2013) and recommended by Logue (2013) and Sturgeon (2006, 2008). A third response, on behalf of
those attempting to provide a positive account of hallucination, is to deny some version of local supervenience—thereby denying that a hallucinatory experience (or a property of such experience) can simultaneously occur in cases of veridical perception (Allen 2015, Moran 2019). A fourth response, on behalf of those attempting to provide a positive account of hallucination, is to deny that explanatory overdetermination is a problem in the relevant cases (Sethi 2020). It seems that the screening off problem is now the most formidable argument from hallucination against naive realism, and so any viable form of the view attempting to provide a general theory of perception must provide a plausible reply.

1.1.1 The Topic and Approach

There are several important questions addressed in this dissertation: What are the core commitments of naive realism? What ontology does naive realism require? What, according to naive realism, determines the phenomenal character of a perceptual experience? What is the intuitive appeal of naive realism? What accounts of the phenomenal character of hallucination are compatible with naive realism? What exactly is the screening-off problem? What naive realist replies to the screening-off problem are viable?

My approach in the dissertation is to bring the tools of analytic philosophy to bear on these questions. In doing so, I take very little of the philosophical ideology used in contemporary philosophy of perception for granted. Instead, I attempt to provide definitions and accounts of many of the key concepts. Second, while my approach has been to follow the argument wherever it leads, I have had to make pragmatic decisions about when to set an issue aside and move on (e.g., I do not weigh in on the debate over
which theory of tropes is best, similarly I do not get into issues concerning causal versus non-causal theories of explanation). Lastly, I am primarily concerned with critically evaluating the most developed, contemporary forms of disjunctivism rather than inventing yet another. Instead, my recommendations for the naive realist will be directional.

There is one more important matter concerning my approach to the topic in this dissertation. Most of the literature on naive realism and disjunctivism is concerned with only the modality of vision and visual sensory experience. Now, I take it that if naive realism is correct for visual experience, the view can be extended, likely with some modifications, to other sense modalities (e.g., gustation, olfaction, etc.) making way for a naive realist account of total perceptual experience. Nevertheless, it remains an open question whether it is more difficult to defend naive realism if we were focusing on other sense modalities or concerned with how perceptual experience is integrated across modalities. A naive realist account of proprioceptive experience may turn out quite different from a naive realist account of visual experience. While my approach throughout has been to drop ‘visual’ modifiers and talk of perceptual experiences more generally, it is more for ease of exposition rather than to assert that naive realism is committed to just the account given across the board for all sensory experience. That said, I do also believe it is worth considering naive realism, at least at times, as a view also about total perceptual experience that includes visual experience. In the spirit of charity, the reader is invited to restrict all unmodified claims concerning experience to visual experience, veridical visual experience, visual hallucination, etc. When the distinction between vision and total perceptual experience is relevant or especially important, I’ll be explicit.
1.1.2 A Road Map

In the rest of this chapter, my aim is to present the core theses of naive realism. Chapter 2 explores the intuitive appeal of naive realism. This is followed by a presentation of two versions of the screening off-problem, evaluating various extant and potential replies, and presenting a novel reply in chapter 3. Finally, in chapter 4, I will present two novel arguments against phenomenal disjunctivism that appear to be fatal to the view.

1.2 The Categories of Perceptual Experience

It is commonplace in the philosophy of perception to distinguish the following three categories of perceptual experience: veridical perception, illusion, and hallucination. In this section, I’ll provide standard descriptions of each while attempting to remain neutral about which philosophical theory of perception is correct. I’ll also introduce two minimal conditions on a philosophical theory of perception.

1.2.1 Three Varieties of Perceptual Success

First, it will be helpful to introduce several lightweight notions of perceptual success. By ‘lightweight’ I mean a concept that is theoretically neutral and without controversial ontological commitment. I’ll introduce the first notion with the help of a familiar construction in natural language. In English, there is a construction with perceptual verbs,  

---

2 I use ‘perceptual experience’ to pick out all experiences of the perceptual variety, including the unsuccessful. I use ‘perceptual experience’ instead of ‘sensory experience’ because the former is more commonly used in the literature on this topic. I will use ‘perception’, ‘visual perception’, ‘veridical perception’, ‘veridical experience’, and ‘veridical visual perception’ to always mean successful cases of veridical perception, which excludes veridical hallucination and veridical illusion. Because of the ambiguity in ordinary English, I’ve made heavy use of ‘veridical’ even though it is cumbersome.
'perceive', 'see', 'smell', 'taste', etc., that take a noun phrase. The basic form of this type of construction is 's perceives <noun phrase>'. The following are examples:

    He perceived a tree.
    Sally sees her dog.
    I smell the sea.

The kind of success encoded in these noun-phrase constructions is that of a perceiver standing in a perceptual relation to an object, generally a mind-independent object in the sensible environment of the perceiver. For now, let's call this relation between the perceiver and the object perceived, perceptual contact. The relation of perceptual contact is theoretically neutral. Depending on one's favored theory of perception, it might be cashed out in terms of causation, counterfactual dependence, a fundamental perceptual relation, or something else.

The first kind of perceptual success, contact-success, occurs when a perceiver, S, is in perceptual contact with an appropriate object in her environment, i.e., S perceives an...
Contact-success is a weak form of perceptual success in the sense that it does not require that the perceived object actually be however it appears to the perceiver. It is also weak in the sense that a subject need only make perceptual contact with a single object for the experience to count as contact-successful.

The second kind of perceptual success involves portraying an object as it is. Let’s call this second kind of perceptual success, appearance-success. Appearance-success occurs whenever a perceiver, S, perceives an object, o, and o is as it appears to S. First, note that appearance-success entails contact-success. Second, while an object can be more or less as it appears, appearance-success is all or nothing. For our purposes here, how an object appears in total to a perceiver sets the threshold for appearance-success. So, for appearance-success to obtain the following condition must be met: for all properties F, if an object, o, appears F to S, then o is F.

---

6 This is a relation that takes perceiving subjects and objects as relata—which is distinct from perceiving its properties. It will become obvious later in this section why this is a non-trivial notion. Basically, in total hallucinations there is no contact success.

7 Sturgeon (2008:112-115) introduces and discusses “contact” and “portrayal”. Note, my treatment of these concepts differs from his in that mine are based on perceiving objects whereas his are based on perceiving worlds. Further, my notion of appearance-success is a success notion, whereas his notion of portrayal is not a success notion. Lastly, Sturgeon thinks you can build a theory-neutral concept of veridicality based on contact determining portrayal measured in degrees. While this idea is elegant, it is difficult to understand without heavyweight supplementation of theory and ontology, thereby undermining its neutrality. I’m not going to go into this in detail here but to get a sense of the challenge, see how much you have to assume to explain how portrayal is not “fully grounded” in contact in cases of Muller-Lyer illusions (or blindsight) over cases of accurate non-illusory conscious seeing.

8 I’m idealizing around a single object. The appearance-success of a total experience requires placing a wide scope quantifier over all objects perceived in the definition: For all objects, o, and all properties F, if o is perceived, then if an object o appears F to S, then o is F. The business of carefully delineating perceptual success gets very complicated. Here, to do this carefully, we will need to restrict what properties are allowed. The first class of needed restriction involves relational properties to other objects. You might have a partial hallucinatory experience of a giraffe on your couch. Intuitively, your experience of the couch is veridical, but the giraffe is hallucinatory. But without restricting the properties in the definition, seeing the couch would not count as appearance-success since the couch appears to have a giraffe sitting on it and there isn’t a giraffe. But the solution isn’t simply to forbid all relational properties involving other appearing objects, otherwise, cases of double-vision will count as appearance-successful. Perhaps, we only allow relational properties to objects with which the perceiver is making contact. I’m not going to pursue this any further here.
Appearance-success does not require the other direction, i.e., that if an object is \( F \), it appears \( F \) to \( S \). If we set the standard that high, then it is likely there would never be a case of appearance-success.\(^9\) Thus, appearance-success is compatible with certain kinds of failure. Suppose if I weren’t sleep-deprived, an object would appear to be some value on a scale but because I am sleep-deprived there is no value that it appears to have. This is a deficiency but quite compatible with our ordinary notions of perceptual success. Culpable omission, my name for this deficiency, is thus a different kind of measure of success or failure--it will not be a focus in what follows.\(^{10}\)

At this point, it may be helpful to make explicit several categories of perceptual failure. First, there is \textit{contact-failure}. Simply put, contact-failure occurs whenever a subject fails to make perceptual contact with any suitable objects--contact failure is total contact failure. Second, there is a straightforward kind of \textit{appearance-failure} that occurs when a subject misperceives a property of an object for which they are in perceptual contact. (Note, for purposes here, both appearance-success and appearance-failure require contact-success.) Moving forward, I’ll reserve the use of ‘appearance-failure’ for this straight-forward kind of failure.

There is another significant kind of perceptual failure in the vicinity of appearance-failure, but distinct from it. I will use the expression, \textit{objectual-appearance}, to pick out an appearance of an object that is neutral with regard to the existence of a perceived object.\(^{11}\) This notion will commit us to “appearances” in a lightweight way without committing to a theory concerning their nature. Let’s call this next kind of perceptual

---

\(^9\) Except perhaps by a supernatural, omniscient being capable of perceiving all.

\(^{10}\) Compare this type of example with writing a report of a battle and omitting through drunkenness or laziness to report on something. This is compatible with everything in the report being correct.

\(^{11}\) Some may prefer ‘to pick out an appearance \textit{as} of an object’ here.
failure, *object-failure*. Object-failure occurs when a subject, $S$, has an objectual-appearance, $\alpha$, and there is no suitable object, $o$, in the environment of $S$, for which $\alpha$ is an appearance of $o$. Next, we can build a corresponding notion of success. *Object-success* occurs when for every objectual-appearance, $\alpha$, that a subject, $S$, has at a specific time, $t$, there is a suitable object, $o$, in the environment of $S$ at $t$ for which $\alpha$ is an appearance of $o$. This is the third lightweight notion of perceptual success. With these notions of perceptual success and failure, I now turn to distinguishing the basic categories of perceptual experience.

1.2.2 Veridical Perception

Veridical perception, as a category of experience, is a perceptual achievement. Veridical perception is standardly defined as an experience where a subject perceives an object and the object is perceived as it is.\(^{12}\) Notice, this standard definition is already problematic because taken at face value it is ambiguous. Further, there is an uncharitable disambiguation (which is how I first hear the standard definition) according to which veridical perception excludes culpable omissions--making it an unapproachable achievement. It is highly unlikely that this super strong notion of veridicality is what philosophers are normally interested in when initially dividing good cases from bad cases. Instead, by using the notions of perceptual success introduced in the previous section (1.2.1), we can disambiguate the standard definition as meaning that veridical perceptions are both object-successful and appearance-successful. This is how I will take the standard definition. Thus, an experience is a veridical perception *simpliciter* if and only if there is

\(^{12}\) Compare Crane and French (2021: section 1.6), Fish (2010: 3), Soteriou (2016: 1).
a suitable object for every objectual-appearance and every way an object appears is a way that object is. Given veridical perceptions are appearance-successful, they are also contact-successful. In short, veridical experiences satisfy all three notions of perceptual success. Next, I’ll present a straightforward example.

As I write this dissertation, I periodically pause and look outside through my office window. I have a visual experience of a tulip tree in the middle of my backyard and the tulip tree looks taller than it is wide. In this example, my visual experience is a veridical perception given that: (1) the appearance of the tulip tree and the appearance of my backyard are both appearances of suitable objects in my environment, (2) I’m seeing a tulip tree in the middle of my backyard, thereby, making perceptual contact with the tulip tree and I’m also seeing my backyard, and (3) the tulip tree I’m seeing is located in the middle of my backyard and the tulip tree is taller than it is wide.

‘Veridical perception’ is a more contemporary term for what is a long-standing category of the good cases of perception. Often, when examining a specific category there are many interesting and subtle subcategories and distinctions to be made. Veridical perception is no exception. There are many interesting distinctions to be made even among the good cases in the lightweight sense. And there will be controversy over what really counts as a “good case”. Take, for example, true belief as a category. There are many interesting subcategories of true belief. When philosophers theorize about propositional knowledge, they often are making fine-grain distinctions between subcategories of true belief. Suppose after spending years doing epistemology in philosophy departments, most philosophers agree to treat true belief as a lightweight notion of knowledge—perhaps as a starting point to which all parties can initially agree. In this hypothetical scenario, all the
philosophers have read Alvin Goldman (2001) on “weak knowledge” and agree to that being one sense of ‘know’. We can make distinctions between various subcategories once we bring in some heavier-weight notions, e.g., justification, safety, sensitivity, defeaters, etc. In my view, we are in a similar, though perhaps less advanced, situation with veridical perception.

One concern with the theory-neutral account of veridical perception I have provided is that it will count some cases of perception that are veridical-by-chance as veridical perceptions. For example, suppose you see a white ball in an undetectable red light. Under normal circumstances, the ball would appear red to you. But in this case, your visual experience is the result of a once-in-a-lifetime encounter by a Cartesian demon who makes the white ball appear white, instead of red, to you. In this case, you see the ball and it is how it appears to you. There is perceptual contact, and it satisfies all the other definitions of lightweight perceptual success; but I suspect many will not want this experience to count as a veridical perception. But in order to exclude such cases systematically, it seems that we would need to bring in heavier-weight commitments to which not all parties would agree. This is like, when doing epistemology, attempting to exclude all the beliefs that are true but insufficiently “connected” to the fact from the good cases of knowledge (where the question of how to spell out the relevant notion of connectedness is up for debate.

---

13 This is an example of a veridical illusion. Intuitively, a veridical illusion is an “illusion” where appearance-success is aberrant or lucky. I’m unsure whether this category can be precisely defined in a lightweight way. The following is another example (that doesn’t involve supernatural interference) from Johnston (2006: 272): Imagine a variation on the Müller-Lyer illusion in which the experiment is constructed using wires suspended in the air instead of lines on a page or screen. We make the wire with the out-arrows longer than the wire with the in-arrows. We place the longer wire on top. Next, we adjust the orientation of the top wire by moving the right side further away from the viewing position. This is not detectable from the viewing position. The result is just what we would expect from the original Müller-Lyer illusion. However, we have two illusions instead of one. The second is the undetected difference in depth orientation. The result is a perceptual experience of the top wire being longer than the bottom wire by just the amount that it is. So, in this way the experience is veridical albeit illusory.
whether we are talking about true belief or genuinely perceiving a fact). One intriguing direction, not to be pursued here, would be to define veridical perception as those perceptual experiences that are absorbed into knowledge (or at least which put one in a position to know by looking). At this point, having stated the concern, I will move forward with the theory-neutral notion of veridical perception as presented.

One other interesting feature of veridical perception worth mentioning here is that it is not undermined by your doxastic stance on the veridicality of your experience. You can undergo a veridical perception even if you believe you are hallucinating and resist being taken in by your experience. Veridical perceptions are contrasted with two categories of perceptual failure: illusion and hallucination. I’ll start with the category of illusion next.

1.2.3 Illusion

An illusion is standardly defined as a perceptual experience which occurs when the subject perceives an object but the object appears other than it is. Don’t be misled by the word ‘illusion’; this is intended to be a notion of misperception, however slight. In the terminology of lightweight perceptual success, an illusion is simply an experience involving an appearance-failure. A classic example of an illusion is the Müller-Lyer optical illusion that involves two equal length lines misperceived to vary in length. In this case, the lines are the same length, but the different arrows cause one line to appear longer than the other line. The line with the arrows pointing inward looks longer than the line with the arrows pointing outward. You may even have confirmed the lines are the same length and nevertheless one line will continue to look longer than the other.
As mentioned, the term ‘illusion’ can be misleading as this broad category is not limited to those tried-and-true classic illusions such as the Müller-Lyer illusion. Illusion is a ubiquitous variety of misperception since it includes all cases where an object is seen but appears to have a property that it does not have. For example, I look out my office window and see a football under the tulip tree, but I don’t see it as a football but rather I see it as a squirrel. The shadows being cast on the football and immediately surrounding grass present a squirrel shape and color. All such cases fit in the perceptual category of illusion. Similarly, when we look at a white wall in red light and have a perceptual experience of a red wall, we are suffering an illusion.

1.2.4 Hallucination

While illusions are perceptual experiences that involve appearance-failure, hallucinations are perceptual experiences that involve object-failure. In cases of hallucination, there is an appearance of one or more objects in the absence of a suitable object or objects (or stimuli) in the environment of the subject. The most familiar cases of hallucination are cases of partial hallucination, where a hallucinatory element is embedded in an otherwise veridically perceived scene. For example, Macbeth has a visual experience of a bloody dagger floating in the air when there is no dagger there to be seen. The rest of Macbeth’s visual experience is assumed to be veridical.

We don’t have to rely on examples from fiction. Hallucinations are a well-documented medical phenomenon. For example, visual hallucinations can be caused by brain lesions, psychotic disorders, somniphany, and can also be drug induced. While visual hallucinations are often partial, they also can be complete in the sense that the total visual
experience does not include any external stimulus. In such cases of total hallucination, there is contact failure. Charles Bonnet Syndrome is a condition which affects some people with severe vision loss as they age in which they suffer vivid visual hallucinations (Ffytche 2013). The hallucinations caused by Charles Bonnet Syndrome are visual hallucinations and are total, rather than partial, in every case where the subject has complete vision loss.

1.2.5 Perfect Hallucination

There are also what are sometimes called *perfect hallucinations* in which a person suffers a total hallucination that is introspectively indistinguishable from a veridical perception.\(^{14}\) Suppose a new drug is discovered that can cause neurological episodes of the visual system from the day before to repeat such that one has the same qualitative type of visual experience regardless of the environment. Let’s call the drug, “ReViz”. Suppose further, I have volunteered to participate in a study to test ReViz. For the study, the drug is administered, and I’m put in a pitch-black room. Next, I have a visual experience of a tulip tree just as I did the day before when having a veridical perception while looking out of my office window. Because I know I’m testing the drug, I’m not taken in by the hallucination. But in a different circumstance, I could have been taken in based on the character of my visual experience. The two visual experiences, one hallucinatory and one a veridical perception from the day before, are so similar that I am unable to discriminate

---

\(^{14}\) Perfect hallucinations have also been called, “philosophers’ hallucinations” (Robinson 2013). Just because perfect visual hallucinations are presumably less likely to occur than other types of visual hallucinations does not mean they are not possible or do not occur. Ffytche (2013) and Penfield and Perot (1963) both show there is empirical evidence that the same neurobiological processes occur in hallucination and matching veridical perceptions. For a dissenting position, see Ratcliffe (2017) who argues that actual hallucinations are not perfect.
between them by introspecting on their character alone. Perfect hallucinations will play a key role in the discussion to follow.

1.2.6 Veridical Hallucination

Veridical hallucinations are hallucinations that depict the world or an object as it is. We can quickly grasp the intuitive idea of a veridical hallucination by reflecting on Lewis’ wizard: A wizard standing before you casts a spell which causes you to hallucinate the scene before you exactly as it is (Lewis 1980: 242). Using the notions of perceptual success, a total veridical hallucination is an experience that is a contact-failure but object-successful.\(^{15}\) Note, total veridical hallucinations are perfect hallucinations.

Next, I will present an example of a total veridical hallucination. I’m sitting in my office looking at the computer screen and I begin to hallucinate the scene of my backyard as it appears when looking through my office window. Suppose this is due to an unknown side effect of ReViz in which it periodically reactivates. This is also a perfect hallucination. Next, I turn and look out my office window. The hallucination is now a veridical hallucination.\(^ {16}\) Suppose next, the drug stops working, the hallucination ceases, and I’m veridically perceiving the scene outside my office window as it is. Further, the transition from hallucination to veridical hallucination to veridical perception is seamless in the sense that there was no variation in the character, at least not any variation introspectively

\(^{15}\) It’s tricky to define partial veridical hallucinations with the lightweight notions—we need to introduce an object appearance relative notion of contact. I could accommodate it, but since this sort of case will not feature in what follows I’m going to skip discussion.

\(^{16}\) It’s not enough for an experience to match some veridical perception of a scene somewhere or other.
detected, of my visual experience over the three experiences. Let’s call this the *seamless transition case*.

### 1.2.7 Minimal Requirements

First, a general philosophical theory of perceptual experience should provide an account of the various categories of perceptual experience. A theory that only provides an account of veridical perception, and has nothing to say about illusion and hallucination, is incomplete. Second, given the possibility of the seamless transition case, a philosophical theory of perception should provide an account of the phenomenal character of visual experience that is compatible with such a case, including the possibility that the various visual experiences are qualitatively “seamless” for the subject. We might think of these two constraints as minimal requirements on a general philosophical theory of perception.

One way to address the second requirement is to say the visual experiences have the same phenomenal character. Before we can explore this in detail, we first need to get clear on phenomenal character. Next, I turn to providing some introductory groundwork on the phenomenality of experience.

---

17 Let’s leave open the exact nature of introspective indistinguishability for now. There is a weak sense of indistinguishability compatible with not introspectively detecting a change in character across the visual experiences in this example. That may be because the character was in fact the same or it remains open whether the character was slightly different but undetected, either undetectable simpliciter or just undetected here but detected somewhere else.

18 Johnson (2004) invokes a similar constraint on a philosophical theory of perception.

19 Presumably, there are more requirements (or adequacy conditions) on a complete philosophical theory of perception. I’m focusing on these as they highlight the importance of providing an account of the character of hallucination.

20 Another way is to say, the character of the experiences are introspectively indiscriminable, though distinct.
1.3 The Phenomenal Character of Experience

The concept of a phenomenal property will play a central role in this chapter. We can get an initial grip on what makes a property a phenomenal property by reference to Nagel’s notion of the subjective character of experience (Nagel 1974). A phenomenal property is any aspect of an experience that characterizes what it is like for the subject to have that experience.

It is tempting to make reference to the what-it-is-like-ness quality of phenomenal consciousness and proceed by way of a few examples. Here, I take a different approach. In this section, I will take the concept of a positive phenomenal property as primitive and introduce it axiomatically. I’ll also take some very general relations like instantiation (between a property and a thing that has it) and undergoing (between an agent of an event and an event) for granted. The payoff will be a precise notion of phenomenal character (axiom 5) and a precise notion of visual experience. I’ll also attempt to precisify several interrelated concepts.

1.3.1 Axioms of Phenomenology

Here are the axioms of phenomenology:

(Axiom 1) Positive phenomenal properties are properties of events.\(^{21}\)

\(^{21}\) I’m not excluding (mental) states from having phenomenal properties, as long as such state talk is understood to be picking out particular tokens, not types. For simplicity of exposition I consistently use ‘event’ talk. I provide some examples of positive phenomenal properties below, phenomenal color and shape are examples. For more on positive on negative see section 1.3.3.
(Axiom 2) An event, \( e \), is an experience, if and only if, \( e \) instantiates at least one positive phenomenal property.\(^{22} \)\(^{23} \)

(Axiom 3) A subject, \( S \), is a conscious being, if and only if, \( S \) undergoes an experience. Any talk of a subject instantiating a positive phenomenal property is to be understood as shorthand for saying that a subject has an experience with a positive phenomenal property.

(Axiom 4) There is something it is like to be a subject, \( S \), if and only if, \( S \) is conscious.

(Axiom 5) The phenomenal character, or simply the character, of an experience, \( e \), is the conjunction of all the positive phenomenal properties of \( e \).

(Axiom 6) If being \( F \) is a positive phenomenal property and being \( G \) is a positive phenomenal property, then being \( F \) and \( G \) is a positive phenomenal property. Positive phenomenal properties are closed under conjunction.\(^{24} \)

(Axiom 7) An experience, \( e \), undergone by \( S \), is total, if and only if, \( e \) has a character, \( C_1 \), such that for all the characters, \( C_2, C_3, \ldots C_n \), instantiated

\(^{22} I'm assuming a fusion of a battle and an experience does not itself have phenomenal properties - in that case we don’t have the bad result that such fusions are themselves experiences. Compare with the following: while I have attitudinal properties, the fusion of me and a rock does not. Now, if one thought that such fusions did have positive phenomenal properties, one would have to learn to live with counting them as experiences or else put tighter demands on what it is to be an experience than I have here. (One candidate I considered: For \( x \) to be an experience it has to have a positive phenomenal property and every part has a phenomenal property. But that is not friendly to naive realists who wish to count a picket fence as part of a veridical experience of a picket fence. I shall not pursue this delicate issue further, especially as the literature is blind to it.)

\(^{23} I remain neutral on whether there is a deep ontological difference between events and objects or more a matter of linguistic artifact (Cf. Sider 2001: 211). Surely experiences, on some level, are more like events than objects. Further, I don’t mean to be taking on any additional commitments in using ‘event’ terminology over say ‘process’, ‘activity’, or ‘episode’ terminology.

\(^{24} There may be some positive phenomenal properties that are incompatible in the sense of it not being possible that the same experience has both at the same time due to facts about a perceptual system or laws of nature.
by all the other experiences contemporaneously undergone by S, having an experience with C₁ entails having experiences with C₂, C₃,...Cₙ.²⁵

(Axiom 8) A subject can have only one total experience.

Since much of the literature focuses on visual experience, it will be helpful to say something about what counts as a visual experience. For this purpose, I’ll take the notion of a *positive visual phenomenal property* as primitive.²⁶

Initially, I was tempted to put forward as an axiom the following simple definition:

An experience e is a visual experience iff e is an event with some positive visual phenomenal properties and which have no other phenomenal properties.²⁷

But *being phenomenally some way or other* is arguably not a visual phenomenal property yet will certainly be had by any experience! Relatedly, suppose *being F* is a visual phenomenal property. Then assuming phenomenal properties are closed under disjunction (which is not an axiom), *being F or being a pain* is a phenomenal property, even though it is arguably not a visual phenomenal property. But if the experience is F, it will be F or a pain. In this case, there won’t be any visual experiences. While I don’t want to rule in disjunctive phenomenal properties, I also don’t want to rule them out if I don’t have to. For both of these reasons, I disregarded this approach.

²⁵ I have intentionally omitted time and duration from the other axioms for simplicity because I don’t think they are needed. Experiences, being particular tokens, have a place and time. They also have a duration. It is fairly straightforward where temporal variables are to be added.

²⁶ I shall not be assuming that if F is a visual phenomenal property and G is a phenomenal property then the property of being F and G is a visual phenomenal property, or being F or G is a visual phenomenal property. I will be assuming that if F is a visual phenomenal property and G is a visual phenomenal property, then being F and G and also being F or G is a visual phenomenal property.

²⁷ To be clear, the idea is that a visual experience is purely visual.
Next, I’ll try another approach to providing a working definition of visual experience, while continuing to take the notion of a positive visual phenomenal property as primitive.

(Axiom 9) An experience, \( e \), is \textit{visually complete} if and only if \( e \) has a character \( C \) such that for every positive visual phenomenal property \( F \) that some experience of yours instantiates, having an experience with \( C \) entails you have an experience with \( F \).

(Axiom 10) An experience, \( e \), is \textit{minimally visually complete} if and only if (i) it is visually complete and (ii) it has no proper part that is visually complete.

(Axiom 11) If a subject, \( S \), undergoes an experience with a positive visual phenomenal property, then \( S \) undergoes exactly one experience, \( e \), that is minimally visually complete. In such cases, \( e \) is called the subject’s “total visual experience”.\(^{28}\)

(Axiom 12) A subject can have only one total visual experience.

Axioms 9-12 attempt to sharpen our notion of a subject’s visual experience by introducing a minimally visually complete experience. Granted some of the axioms can be challenged. But as they are implicit within most of the literature on this topic, I shall largely assume them myself in this work. The foregoing helps things along not by articulating what is

\(^{28}\) I’m not assuming that the minimally visually complete experience has only positive visual phenomenal properties as its positive phenomenal properties. As before, we might want to count \textit{being phenomenally some way or other} as a positive phenomenal property but not a visual one.
completely obvious but instead by making explicit what is largely inchoate and implicit in much of the literature. Next, some commentary on the axioms presented above.

1.3.2 Commentary on the Axioms

First, I have made experiences, understood as events, the primary bearers of phenomenal properties. One might disagree with this general approach. Some may find this an affront to their ontological sensibilities on the grounds that it is more intuitive that subjects, rather than experiences, instantiate phenomenal properties. Suppose I were to suddenly explode in an accident. Presumably, the property of exploding and myself are in some sense more “fundamental” than the explosion event, since the event is arguably “constructed” out of the other entities: myself qua subject and the property of exploding. Being an explosion (a property of an event) is arguably a less fundamental property than the property of exploding (a property that I and not the event instantiate).

In the debate over naive realism, it’s the notion of the ‘character of experience’ that is front and center. Perhaps philosophers have talked about this subject in a manner that is ontologically misleading; but their surface grammar often attributes phenomenal properties to experience (e.g., ‘qualitative character of experience’, ‘phenomenology of visual experience’, ‘a painful experience’). Regardless of parlance, it seems that there are three critical components: subjects, experiences, and phenomenal characters. I have chosen to follow Williamson (2013: 48) in developing the framework as I do; subjects have experiences and experiences are the primary bearers of phenomenal properties and characters. Recently, Papineau (2021: 14) departs from the Williamsonian framework and makes a point of starting with people and their phenomenal properties as fundamental.
Also, Pautz (2010) in his comprehensive discussion of the phenomenology of visual experience employs the notion of a ‘visual experience property’, which is instantiated by subjects.\textsuperscript{29} Finally, Chalmers (2010: 341) is content to switch back and forth between attributing phenomenal properties to mental states and people.

Given axioms 1-3 above, it is easy enough to switch back and forth between attributing phenomenal properties to experiences and to subjects. We can use the following biconditional as a translation guide:

A subject, \(S\), has an experience, \(e\), which instantiates the phenomenal property \(P\), if and only if, \(S\) instantiates the property of having a \(P\)-experience.

When I take off my glasses and look out my office window, I have a visual experience that instantiates the property of being blurry. I also instantiate the property of having a blurry visual experience. Technically, the property having a blurry visual experience is not a phenomenal property, as I have defined it.\textsuperscript{30} But its instantiation does entail the existence of a particular experience and the instantiation of a phenomenal property, in this case, being blurry. We can tolerate loose talk calling both formulations “phenomenal properties” as long as we are clear about what is going on.

There are other uses of the expression ‘phenomenal property’ in the vicinity that need to be flagged as to avoid confusion. In particular, Dretske (1995, 2000) and Tye

\textsuperscript{29} Note, this is a slightly different kind of property. Pautz’s property is basically the ‘P-visual experience property’ presented in the next paragraph. I show how we can move back and forth between phenomenal properties instantiated by experience and a related property instantiated by people.

\textsuperscript{30} To be clear, this is why having a translation guide is not the same as showing there is no genuine or deep disagreement about what property is more fundamental. Moreover, just as there is arguably a distinction between being the agent of an explosion event and exploding, there may be a distinction between being the agent of an experience that is blurry and some other property that stands to that property as exploding stands to being the agent of an explosion. I shall not be exploring these delicate issues further. They turn in part on whether we should be “fine grained” about properties, countenancing distinct pairs of properties that are necessarily coextensive.
(2009, 2015) hold that a ‘phenomenal property’ is a property of an object represented in experience. But given axiom 3, it would be a category mistake to attribute phenomenal properties to objects. When you are in a park looking at the lawn, you have an experience of greenness, where greenness is a property of the lawn. Your visual experience is greenly, where greenly is a phenomenal property of the experience (or at any rate the character of your experience entails you have an experience that is greenly).31 The conceptual distinction between the sensible properties of objects, represented (or presented) in experience, and the phenomenal properties of experience is critically important. Of course, one might hold that to be greenly is to represent greenness. But that would be to advance a substantive thesis about the nature of perceptual experience.

Given there is something it is like to be you in the park looking at the lawn, axiom 4 implies you are conscious. Given you are conscious, axiom 3 implies you are having an experience. Given you are having an experience, axiom 2 implies your experience has a positive phenomenal property. In this case, presumably your experiences instantiate many phenomenal properties. One such phenomenal property is greenly. Greenly is one aspect of your visual experience. This aspect of your visual experience changes when you look straight up to the sky on a clear day. When you look up, your visual experience is bluely. In this example, I have pointed out the differences in phenomenal color properties between two sequential experiences. So, the two experiences differ in at least one positive phenomenal property. The following biconditional is a corollary of Axiom 5:

31 In the main text below, I explain how the ascription of positive phenomenal properties to experiences, both total and minimally visually complete, is a delicate matter and my general approach.
Two experiences, $e_1$ and $e_2$, differ in character, if and only if, $e_1$ and $e_2$ differ in one or more positive phenomenal properties.

So, your experience when looking at the grass and your experience when looking up at the sky differ in character.

Accounting for similarity and difference in character is a key topic of this chapter.

The question is often put this way:

What is it to have a perceptual experience with a certain character?

I’ll call this the character question.32 We will look at particular answers to the character question below. But first, back to clarifying the remaining axioms.

1.3.3 Positive Phenomenal Properties

Axioms 1, 5, and 6 employ the notion of a positive phenomenal property. Now, I am tempted to think that a property is a phenomenal property if and only if it is a positive phenomenal property. I’m not saying I have anything against negative properties.33 Assuming that being a pain is a positive phenomenal property, I have no issue with postulating the property not being a pain. But it seems a bit strange to count not being in pain as a phenomenal property. If we are counting the negation of any phenomenal

---

32 I’m following Pautz (2021: 4) in calling this “the character question” but he proceeds by way of experience-types. Basically, different experiences are typed by their character.

33 Paolini Paoletti (2014, 2017) argues for negative properties. Armstrong (1979: 23-29) makes a strong case against negative properties with several different arguments, though admits they are not definitive. The most powerful, albeit pragmatic, of Armstrong’s arguments is that we have no need to postulate negative properties and negative states of affairs to provide truth conditions for statements with negative predicates. There are several methods for doing this. He sketches a method involving counter-correspondence attributed to Keith Campbell. Roughly, the idea is that ‘Pa’ counter-corresponds to Ra, Qa, etc for all of a’s properties. Then ‘~Pa’ corresponds to the world iff for each property, $\varphi$ such that $\varphi a$, ‘Pa’ counter-corresponds to $\varphi a$. 
property as a phenomenal property, we would have to say that everything in reality has the same number of phenomenal properties. Armstrong (1979: 24) held that the result that every particular must have the same number of properties “should appall any empiricist”. Anyone with any sympathy to Armstrong’s complaint will think that the view that every particular must have the same number of phenomenal properties is especially appalling!

More seriously, negations of positive phenomenal properties seem “less phenomenal” in nature. The theoretical role of ‘phenomenal property’ in the literature seems to be that they are indicative of consciousness. But negations of positive phenomenal properties are obviously not so indicative. So, negations of phenomenal properties are not a serious challenge to the view that a property is phenomenal if and only if it is a positive phenomenal property. But there are other tricky issues. Take the disjunctive property of being a pain or a dog. I clearly do not wish to count such a disjunction as a positive phenomenal property (since positive phenomenal properties can only be had by experiences). Since I didn’t want to get into the issue of whether to count such properties as phenomenal, I skirt the issue by using ‘positive phenomenal’ in the axioms. By framing the thesis for consideration in terms of positive phenomenal properties, the question as to whether that disjunctive property counts as a phenomenal property becomes irrelevant to the work ahead.\footnote{Similarly, while I don’t want to count the conjunctive property of being a pain and near a giraffe as a positive phenomenal property (for one thing, that would prohibit experiences ever having matching characters), I can regard the question as to whether that counts as a phenomenal property as irrelevant for current purposes.}
1.3.4 Total Experience and Visual Experience

Axiom 7 spells out the relationship between the character of the total experience and the character of other experiences, e.g., visual experience. The character of your total experience as you read this sentence entails the character of your visual experience over the same period of time. While it would be elegant if all of the positive phenomenal properties of your visual experience were inherited by your total experience, such a principle of inheritance fails for positive phenomenal properties. Your total experience is not blurry whenever your visual experience is blurry.

There are subtle issues here. The care taken in formulating a notion of a minimally visually complete experience was partly motivated by the constraint that the bottom-up inheritance of positive phenomenal properties between experiences fails.35 Returning to the park example. Let's start with the property of greenness that is attributed to the lawn. Are we talking about being green all over so that a lawn that is only half green is not in the relevant sense green? Now a similar issue arises for greenly. If the right half of my visual field is pervaded by a thing that is green all over and the left half is pervaded by a thing that is blue all over, should we say that the visual experience (in my sense) is greenly or merely that the experience corresponding to the right half of the visual field is greenly? If greenly stands to experience as “green all over” stands to objects, then it is wrong to say that the visual experience is greenly. Of course, the visual experience can still perfectly well be minimally visually complete--since, even if it does not instantiate being greenly, it instantiates a character that entails that some experience is greenly. Again, it is tempting

35 By ‘bottom-up’ I mean an experience, e, inherits positive phenomenal properties from other experiences that are experiential parts of e, e.g., your visual experience is bigger than your left sided visual experience, or your total experience is bigger than your visual experience.
to assume a bottom-up inheritance principle such that whenever an experience instantiates a positive phenomenal property, then any larger experience it is a part of will inherit that property. But as this example makes clear, that assumption is very tendentious indeed. That said, I shall often myself be casual about such issues, though we should be on the lookout for places where this kind of issue really matters to the argumentation.

1.3.5 On Character

If we combine the axioms with the assumption that events are not the kind of thing that repeat, it follows that phenomenal properties and characters are *types* and experiences are non-repeatable tokens. But notice that it may be that two experiences $e_1$ and $e_2$ instantiate some property *being F* but that *being F* is only the character of $e_1$. Suppose, as a simple toy example, that $e_1$ had exactly one positive phenomenal property, *being F*, and that $e_2$ has three, *being F*, *being G* and *being F and G*. Then, *being F* is the character of $e_1$, and *being F* is instantiated by $e_2$, but *being F* is not the character of $e_2$. In short, the

---

36 I think we can get duped into thinking inheritance holds because total experience has visual experience as a proper part. But it doesn’t follow that total character has visual characters as a part as defined. Just as tables don’t inherit all the properties of its legs, e.g., size. I suspect this is likely a consequence of the general trend to use ‘phenomenal character’ talk loosely to mean any and all of the following: a positive phenomenal property, some conjunction of positive phenomenal properties, all the positive phenomenal properties of total experience, or the positive phenomenal properties of some experiential part of total experience. By ‘experiential part’ I mean a part that is also an experience, e.g., visual experience. Again, this is part of the motivation for formulating the axioms in the first place.

37 Where I sloppily say the visual experience is *F-ly*, the reader can treat this as ‘*F-ly*’ where to be *F-ly* is to instantiate that disjunction of characters that entail having some experience that is *F*.

38 I’m following Williamson (2013) in treating phenomenal characters as (1) maximally specific and (2) types of which experiences are unrepeatable tokens. But my usage differs in not committing to phenomenal character even being weakly subjective, that is when indiscriminability and identity coincide in the sense that characters are distinct only if there is at least one pair of presentations that are discernable. Otherwise, the question is begged against the disjunctivists at the outset. Martin (2004, 2006) holds that two experiences can be indiscriminable, in the sense that it is not possible to know they are distinct on the basis of reflection alone, even though they differ in character. So, in such a case I understand Martin as holding that there is no possible pair of presentations where they are discriminable, at least on the basis of only reflecting on the experiences.
property of being a character, *if instantiated*, may be a contingent property of some conjunctions of positive phenomenal properties.

The axioms of phenomenology, as presented above, are silent on a number of controversial topics. Most importantly, the axioms are intended to be compatible with various more substantive accounts concerning the nature of these properties, e.g., whether phenomenal properties are determined by representational content or directly by the properties of mind-independent objects. Further, I see no reason here to take a stand on whether our knowledge of phenomenal properties is *infallible* or if they are, in some sense or other, *ineffable*.

The axioms are also silent on the relationship between any pair of introspectively indiscriminable experiences and their characters. For example, take the following principle:

\[(\text{IND}) \text{ If two experiences, } e_1 \text{ and } e_2 \text{, are indiscriminable for their subject,} \]
\[\text{then } e_1 \text{ and } e_2 \text{ have the same character.} \]

On standard interpretations of its key terms, (IND) is incompatible with most versions of naive realism since the character is determined by objects and properties perceived and *prima facie* there are indiscriminable experiences which differ in their objects. For now, note that the axioms are compatible with (IND) and its negation. In the next section, I turn to naive realism.
1.4 Naive Realism

*Naive realism* is a philosophical theory about the metaphysical nature of veridical perception.\(^{39}\) In this section, I introduce and discuss the core theses of naive realism (1.4.1). In order to clarify key concepts, I present an overview of fundamentality (1.4.2) and discuss various subtleties concerning the directness of veridical perception (1.4.3).

1.4.1 The Core Theses of Naive Realism

While the label, “naive realism”, sometimes involves other commitments, I mean to pick out the class of theories committed to the following four theses:

**Directness Thesis:** When a subject stands in the relation of *perceptual awareness* to mind-independent concrete entities, that is not (even in part) in virtue of standing in the perceptual awareness relation to entities that are not mind-independent concrete entities, e.g., sense-data, mental images, propositions, etc.

**Ontological Thesis:** In veridical perception, a subject stands in the relation of *perceptual awareness* to mind-independent entities, and this relation is the most fundamental of all the perceptual properties and relations.

---

\(^{39}\) Any naive realist account of veridical perception is a further refinement of the theory-neutral account of veridical perception provided in section 1.2.2. The lightweight notion and theirs will likely come apart in the sense of disagreeing over whether a particular case is a good case or not. I’m not going to distinguish lightweight versus heavyweight versions everytime I use ‘veridical perception’ as the context will make clear what is going on.
Phenomenological Thesis: In veridical perception, the character of experience is partly determined by the particular mind-independent entities to which the subject stands in *perceptual awareness*.

Externality Thesis: In veridical perception, mind-independent entities are constituents of the event that is the perceptual experience.

I’m using ‘mind-independent entities’ in the theses above in an expansive way to include objects, properties, tropes, events, etc. The dominant approach to developing naive realism is one according to which the kinds of mind-independent entities that we stand in the perceptual awareness relation to include trees, tables, dogs, pencils and so on. But that is not mandatory. (If for some reason you are desperate to use ‘naive realism’ in a way that entails that some people sometimes see fences, then you can easily supplement the core theses above with a requirement that we stand in the perceptual awareness relation to ordinary physical objects.) For all I have said, it may be that we never stand in the perceptual awareness relation to dogs or tables.⁴⁰ Other candidate relata include (i) dog surfaces (ii) tropes, but never their bearers (iii) events befalling light (iv) bits of the subject’s brain. That said, I shall, except where explicitly noted, tend to assume as a default that the naive realist will include ordinary physical objects such as chairs and dogs among the relata of the awareness relation.

---

⁴⁰ If we never stand in the relation of perceptual awareness to ordinary objects, does the view still count as “naive” realism? It isn’t clear why we call this view naive realism anymore. The views are quite complex and far from naive. In any case, I don’t care about what we call the view. Perhaps ‘direct realism’ is better. In the literature on the metaphysics of experience, the view captured with the core theses is most often called ‘naive realism’.
Contemporary naive realists include Allen (2020), Brewer (2011), Campbell (2002), Fish (2009), Martin (1997), Moran (2019), and Langsam (2017). They each use slightly different terminology but defend a view committed to (at least something in the vicinity) of the four theses above. They also each go beyond these minimal commitments in various ways which distinguish their particular version of naive realism from others. However, the theses above capture an important core of naive realism for the discussion to follow. Next, I will examine and clarify each thesis in turn.

1.4.2 Fundamentality

The most important component of the ontological thesis in need of clarification concerns the nature of the relation of perceptual awareness. This special relation is described by proponents of naive realism as being “primitive”, “fundamental”, “irreducible”, “ineluctable”, and “unanalyzable”. Further, sometimes those terms appear to describe a relation as a part of reality and other times it seems to be describing our concept of the relation. To be clear, the ontological thesis is about the nature of the perceptual awareness relation as a part of reality. I have deployed the expression, ‘most fundamental’, in the ontological thesis which I’m using to

41 I was tempted to list Alston (1999) and Johnston (2004) here but each of their views is different enough from the naive realism of the others that they are each better considered (direct realist) close cousins. Here, I have referenced one representative book or paper for each philosopher. Each person cited has more than one publication arguing for naive realism and other papers for each can be found in the references section. There are more philosophers who are defending naive realism (if not outright endorsing it): French and Gomes (2019), Hellie (2013), Logue (2012A), Nudds (2013), Raleigh (2014), Soteriou (2005).

42 I say this because ‘primitive’ and ‘unanalyzable’ are often properties of concepts, while ‘fundamental’ and ‘irreducible’ are often used as properties of properties and relations.
describe the special status of this relation among other perceptual properties and relations. Next, I’ll offer a rough account of how I’m using that expression.

My characterization of the perceptual awareness relation relies on a graded notion of *fundamentality* for properties and relations. Within the domain of conscious perception, there are many properties and relations that are of particular philosophical and scientific interest such as: perceptual properties of events, relations between subjects and intentional contents, relations that hold only between subjects and mind-dependent entities (such as sense-data or mental images), and the relation of perceptual awareness. The ontological thesis need not deny the existence of any of these perceptual properties and relations. It merely insists that they are less fundamental than the special relation of perceptual awareness.

What do I mean by ‘fundamental’? To invoke a slogan, the most fundamental properties and relations “carve reality at its joints”. Lewis (1983) calls the properties and relations that carve reality at its joints, the *natural* properties and relations. I invoke Lewis’ notion of naturalness in elucidating fundamentality, at least for current purposes, because it is a well-developed notion with many applications and a rich subsequent literature. Further, Lewisian naturalness is objective and also comes in degrees. So by ‘fundamental’ and its cognates, I will mean ‘natural’ in Lewis’ sense. According to fundamentality enthusiasts, we can provide an exhaustive description of the world by providing a complete distribution of the most fundamental properties and relations. All the other properties

---

43 See Dorr (2019), Dorr and Hawthorne (2013) and their bibliographies. Also, compare Sider’s account of structure as a theory of fundamentality (2011: Chapters 1 and 7).

44 I’m using Lewis and his account of naturalness as a standard bearer of a theory of fundamentality in metaphysics. But I’m not committing naive realists to the details of Lewis’ account. The reader may plug in their favorite theory of fundamentality instead, as long as it is up for the task outlined in the text.

45 Lewis (1983, 1986) calls these the *perfectly natural* properties.
and relations are built out of the most fundamental properties and relations. Properties and 
relations can be more or less fundamental than other properties and relations. Note that 
my formulation of naive realism is compatible with the thesis that perceptual awareness is 
one of the most fundamental relations *simpliciter*. But I don’t want to commit them to 
anything so strong.

Further, I shall also assume that the more fundamental is prior in order of 
explanation to the less fundamental. Thus, the account of fundamentality used here gives 
rise to claims along the following lines: perceptual awareness relations are more 
*explanatory* than other perceptual properties and relations. Of course, it is conceivable that 
one goes in for a more complex metaphysics where there is a mismatch between the order 
of explanation and the order of fundamentality but exploring that issue further would take 
us too far afield. Let’s not get sidetracked in metametaphysics any longer than necessary. 
The foregoing is enough to understand my conception, in broad strokes, of what it is for a 
property or relation to be more or less fundamental than other properties and relations, and 
why I am not using ‘explanatoriness’ as a separate moving part.

Next, a few words on *reduction* are in order. The literature I am exploring often 
uses this ideology without explaining it and without being fully clear as to what the 
relata of reduction are. To fix ideas, I will take it as a relation between families of 
properties and relations. Let us say family $A$ is reducible to family $B$ if and only if (i) 
everything in $B$ is more fundamental than everything in $A$ and (ii) there is a language $L$ 
in which the only non-logical expressions express the members of $B$ such that, for every 
member of $A$, a property intensionally equivalent to $A$ can be expressed by some
From the ontological thesis, it follows that there is no reduction available of the relation of perceptual awareness to either a family of other perceptual properties and relations or a family of other perceptual properties and relation plus some other properties and relations—since, if perceptual awareness is the most fundamental of the perceptual properties and relations, no such family can pass the reduction test.

A few observations about this framework of reduction: (1) no property reduces to itself, since the family containing some property \( F \) cannot be more fundamental than the family containing that very property, (2) this is all perfectly compatible with the idea that ‘\( x \) is perceptually aware of \( y \)’ is expressible by other, more complicated, psychological predicates, e.g., ‘(\( x \) is perceptually aware of \( y \) and \( x \) is in pain) or (\( x \) is perceptually aware of \( y \) and \( x \) is not in pain)’ obviously expresses a relation that is intensionally equivalent to the relation of perceptual awareness but does not point to a reduction in my sense, and (3) it doesn’t follow that perceptual awareness doesn’t reduce to some family of properties drawn from biology plus neurology plus causation, etc. To show perceptual awareness is irreducible to such a family of properties and relations, the naive realist will need to provide supplemental argumentation.

1.4.3 Directness

Next, I make a few clarificatory remarks regarding the directness thesis. As I have previously stated, naive realism, as I have presented it, need not deny that we stand in some

\[^{46}\text{If you think intensional equivalence between properties suffices for identity, then you can replace ‘intensionally equivalent to’ with ‘identical to’. I do not wish to get enmeshed in matters of grain science for properties here, i.e., questions about how fine-grained properties are.}\]
kind of perceptual relation to propositions, nor that we stand in perceptual relations to sense data or mental images.\textsuperscript{47} It doesn’t even follow from the core theses of naive realism that we can never stand in the haloed perceptual awareness relation to, say mental images.\textsuperscript{48} All the naive realist will insist on is that we do not stand in that special relation to mind-independent concrete entities (e.g., external physical objects) in virtue of standing in that relations to entities that are not mind-independent concrete entities (e.g., internal mental objects).\textsuperscript{49} Now of course there is a version of naive realism that insists that we never stand in the special perceptual awareness relation to internal mental objects. That version might deny the existence of a certain class of mental objects. On another version, they may concede their existence but think that there are other perceptual relations, apart from the special perceptual awareness relation, that we stand to internal mental objects. Nothing in naive realism rules that out. What is crucial of course is that the special perceptual awareness relation is not reducible to, say, a family that includes causation plus various other perceptual properties and relations. That follows, as noted above, from the ontological thesis.

1.5 Veridical Phenomenology

The phenomenological thesis provides the naive realist answer to the character question: What is it to have a perceptual experience with a certain character? To be clear, it only

\textsuperscript{47} A naive realist might deny there is any such perceptual relation to propositions. But it is obvious that we stand in a relation to propositions when we perceive things. So, the central quibble will be about whether to count such relations as “perceptual”. I would prefer not to get too hung up on delimiting the rather vague category of ‘perceptual properties and relations’.

\textsuperscript{48} Alston (1999) entertains such a view in cases of hallucination.

\textsuperscript{49} It doesn’t even strictly follow that we never stand in the special perceptual awareness relation to propositions as well as, say, dogs, though I take it that it is rather implausible to suppose that one and the same relation can do the job here.
provides an answer to the character question for veridical perceptual experiences. Below, I will be exploring what answers to the character question are viable for the naive realist in cases of hallucination. In this section, I’ll attempt to clarify the phenomenological thesis.

1.5.1 Campbell and Martin on Character and Constituency

It may be helpful to start by reviewing how two prominent naive realists express the thesis in their own way. Here is an often cited passage from Michael Martin:

According to naive realism, the actual objects of perception, the external things such as trees, tables and rainbows, which one can perceive, and the properties which they can manifest to one when perceived, partly constitute one’s conscious experience, and hence determine the phenomenal character of one’s experience. This talk of constitution and determination should be taken literally; and a consequence of it is that one could not be having the very experience one has, were the objects perceived not to exist, or were they to lack the features they are perceived to have. (1997: 83)

What stands out in the passage is that both the phenomenological thesis and the externality thesis are intertwined in Martin’s presentation. And that is not uncommon. Just below I’ll explain why I find it helpful to clearly delineate the two theses. But first, here is an often cited passage by John Campbell:

[According to naive realism]...the phenomenal character of your experience, as you look around the room, is constituted by the actual layout of the room itself: which particular objects are there, their intrinsic properties, such as colour and shape, and how they are arranged in relation to one another and to you. ...two ordinary observers standing in roughly the same place, looking at the same scene, are bound to have experiences with the same phenomenal character. (2002: 116)
Note that the notion of constituency appears to be doing some work in how both of these naive realists understand how the perceived entities determine character. But it is not mandatory.

The account of phenomenal properties and character based on the axioms of phenomenology (1.3) is compatible with the truth or falsity of the externality thesis. Both of these naive realists are expressing the basic idea that the constituents of an experience partly determine the phenomenal character of that experience. Further, it is implied that any entity to which the subject stands in the perceptual awareness relation is a constituent of the experience. The theses fit together, and there may be something highly intuitive about the fit, but I don’t see how the externality thesis alone deepens our understanding of the phenomenological thesis. Hence, my choice to separate the phenomenological thesis and the externality thesis.

1.5.2 Two Questions about the Determination Relation

Next, I’ll attempt to unpack the relation of partial determination as used in the phenomenological thesis. In what follows, I’m going to use the expression, ‘entities of awareness’, as an abbreviation for the expression, ‘the particular mind-independent entities to which the subject stands in perceptual awareness’ as used in the phenomenological thesis. Concerning ‘partial determination’, two questions linger:

(1) If the entities of awareness only partially determine character, what else is involved?
(2) What is the nature of the determination relation as used in the phenomenological thesis?

Next, I’ll take each question in turn.

1.5.3 Edenic Realism and Perceptual Variation

Let’s start to answer the first question by initially considering a super naive variant of naive realism, edenic realism. According to edenic realism, the character of experience is completely determined by the entities of awareness. Consider the following edenic slogan about the character of veridical perception:

(M) A match in entities of awareness, guarantees a match in character.

The term ‘match’ can have many different meanings, but here in this section entities of awareness match if and only if they are identical.

Edenic realism faces counterexamples. The counterexamples involve variations in character when perceiving the same object (or entity) between different subjects or the same subject under different circumstances. Further, in such cases we are hesitant to count any of the variations as a non-veridical case. Let’s look at three kinds of counterexamples. The first kind of counterexample is based on a personal anecdote. There are differences in the character of visual experiences I undergo when viewing the exact same screen of my phone (from the same position and distance, looking in the same

50 For further discussion of such a view, see Chalmers (2006). The theory of perception, what I’m calling ‘edenic realism’, is true in Chalmers’ edenic worlds.

51 The debate here is familiar from the philosophy of color. The arguments from perceptual variation in veridical cases are levied against color physicalism and in favor of color relationism. See Allen (2016: Chapter 3) and Cohen (2009: Chapter 1 and 2). I’m going to rely on the same arguments here.
direction, in the same lighting conditions both the screen setting and the lighting in my office) at different times of day and also on different days. Basically, there are times when I can see what is on the screen more clearly than other times, likely based on the condition of my eyes. But I don’t misperceive the screen of the phone in either case. This kind of counterexample is a case of intrasubjective variation.

A second kind of example involves the robust empirical finding that there is wide variation in the experience of phenomenal color, even among people with normal color vision under the same conditions. Hardin (2004) explains the findings with a study involving unique green. A color looks unique green just in case it looks green without at all looking bluish nor at all looking yellowish. When asked if a specific Munsell chip, G, is unique green, there is variation between normal subjects under the same viewing conditions: some subjects report G is unique green, while other subjects report G is not unique green. This kind of counterexample is a case of intersubjective variation. These empirical results of psychophysical variation across normal subjects are robust. Further, there is no good reason to favor one chip as being the veridical case of looking unique green.

There are also empirical results from a similar study design that further support intrasubjective variation (Hurvich, Jameson, and Cohen 1968). The same subject will vary in reporting which Munsell chip looks unique green when the following viewing conditions are altered: illumination, background color, viewing distance, viewing angle, etc. Given there is no principled reason on which to count a set of viewing conditions as “ideal”, it is argued that there is no good reason to hold that a veridical perception of unique green is
one specific Munsell chip looking unique green over other Munsell chips under different conditions.

There is a third kind of counterexample regarding a difference in perceptual systems. Humans have trichromatic color vision that involves three channels of information. Two channels involve color opponent processing and a third channel both for luminance and black and white processing. Dogs have dichromatic vision with just two channels for optical information (Neitz, Geist, and Jacobs 1989). Empirical research in this area supports that (1) dogs see considerably less colors than humans (10,000 versus 1,000,000), and (2) dogs do not experience color in the same way and don’t experience phenomenal redness. Thus, when looking at the same red ball, a human and a dog are undergoing an experience with a different character. This kind of counterexample is an example of interspecies variation. There doesn’t seem to be a good reason to count either humans or dogs as the species in perceptual error given there is a clear sense in which both are veridically perceiving given their respective perceptual systems.

These three kinds of counterexamples show that a match in the entities of awareness, understood as a mind-independent object, does not guarantee a match in character. Now, a defender of edenic realism can still put up a fight here. One might argue that the entities of awareness are not the same in each of the putative counterexamples. There is one object, o, in each example, but each object, o, has many properties. The “properties” of which the subjects are perceptually aware are different across the cases in each example and explain the variation in the character of each experience.52 Next, I need

52 Of course, it may be important for the naive realist not to be too naive. The completely flatfooted approach might say this: whenever we say we saw a dress as swanky, that is because we stand in the fundamental perceptual relation to the property of swankiness. Even if there is such a property as being swanky, it seems implausible that the perceptual awareness is reaching out to instances of swankiness. At
1.5.4 Particularity, Properties, and Tropes

According to naive realism, we perceive particulars in veridical perception. Both Martin and Campbell, in the quotes above, talk of subjects perceiving ordinary objects in the environment and the properties of those objects. It is important to appreciate that the properties perceived in veridical perception are properties of particular objects in the environment--and, as such, are instances (tokens) not universals (types). We see the color of a particular object, not color in general: So, for example, I don’t see the universal property, being red. I see a red property-instance. I’ll call a property-instance, a trope.

Tropes are particulars. A concrete object may have many tropes. When naive realists talk of perceiving objects and their properties, the reference to properties is always to particular property-instances, that is, tropes. I take it that no naive realist contends we automatically perceive the properties of an object simply in virtue of standing in the perceptual awareness.

---

53 I will not be entering further debate over different trope theories. Tropes, neutrally understood as property-instances are compatible with different overarching ontological theories. Obviously, I’m not saying that naive realists are committed to some radical ontology whereby the entire world is built solely out of tropes. But I am saying that naive realism requires tropes (perhaps among other basic categories). One issue I won’t get into is whether, when I see a dark red thing, the same trope somehow manifests both the property of being red and the property of being dark red, or whether instead there are different tropes corresponding to the more and less specific universals. Note that the expression ‘being an instance’ does not quite line up with the ordinary use of ‘instance’ for properties. Normally, when we look for an instance of chairhood, we point to a particular chair. The concrete chair instantiates being a chair and so is an instance of being a chair. Similarly, you might think an English letter box is an instance of being red. Trope theorists need a different notion of instance. For an overview on tropes, see Maurin (2018).
relation to objects alone. So, in order to explain variations in character based on variations in which properties of objects we perceive, naive realism requires that subjects are perceptually aware of tropes. Next, I’ll present additional arguments to further substantiate that claim.

If we are defending naive realism, why not just say that the relation of perceptual awareness is always to ordinary objects and never to tropes? The basic thought is that the character of the experience is determined by which things stand at the end of the fundamental relation of perceptual awareness. But it is difficult to grasp how this will work if the relata are limited to objects. Take two cases, one case where I see Ernie and Ernie is red and another case where I see Ernie and Ernie is blue. If the only thing that I am perceptually aware of is Ernie in each case, then the two cases match with regard to which object is at the end of the perceptual awareness relation. But the character of my experience is very different in the two cases. We no longer have a picture where the character of our experience is determined by what is at the end of the perceptual awareness relation.\(^{54}\)

There is a second option. Why not deny that we see tropes and try to do all the work with (i) objects like Ernie and (ii) universals? Here are some reasons for concern: First, it is natural to think we see particulars, and universals are not particulars but tropes are.\(^{55}\) Second, suppose I see Ted and I also see Dean. And I see Ted is kicking Dean in a way that catches my attention. This is not easily accounted for simply by saying: I see Ted, I see Dean and I see kicking. For even if Dean was kicking Ted, I would presumably still

\(^{54}\) Obviously, any other variables (standpoint) that determine character are held fixed across the cases, other than the color of Ernie.

\(^{55}\) If you have been taken in by Armstrong (1997), you might think there are (a) thin particulars that are in the ballpark of the Lockean “bare” particulars and (b) “thick” particulars. Armed with that metaphysics, you might want to ask whether we ever see thin particulars or just thick particulars? Since I don’t believe in thin particulars--and neither, as far as I can see, do other contemporary metaphysicians--I think the question about thin particulars is a non-issue.
be seeing kicking. Somehow the kicking relation has to be bound in the right way in the experience. Call this the binding problem.

Similarly, if Ted is blue and Dean is red you might not think the character is accounted for by seeing each of Ted, blue, Dean, and red, since they could be the other way around. Consider also the case where Ted is a red square and Dean is a blue circle. The character of the experience cannot be accounted for by saying that we see redness, blueness, squareness, circularity, Ted, and Dean, because we would see all those even if Ted was a blue square and Dean a red circle. Now, if one had propositions at the end of the fundamental perceptual relation there would be no need for tropes here. But if we do not want a relation to a proposition to determine character, the turn to tropes is very tempting. The counterintuitive alternative is to say (1) that the character is no different in the case where I see Ted kicking Dean and the case where I see Dean kicking Ted and (ii) that the obvious difference in character between the case in which Ted is a red square (and Dean is a blue circle) and the other case in which Ted is a blue square (and Dean is a red circle) can be accounted for by saying that we see complex universals. In the one case, we see two conjunctive universals, being blue and circular, and being red and square, but not in the other.

There is another reason naive realists want to talk about seeing a red square as a red square. Why should seeing Ted and seeing that conjunctive property (i.e., being red and square) suffice for seeing Ted as a red square? Note here that the official ideology posits an awareness relation between subjects and particulars. The three-place relation a
subject, $S$, seeing $x$ as an $F$ (or $S$ experiences an $F$ as an $F$) requires further analysis. Tropes seem to provide the best shot.\textsuperscript{56}

Some naive realists hold that in veridical perception we stand in perceptual relations to entities that are of the form: $x$ is $F$. Now on the face of it this is a capitulation to a kind of intentionalism, since the paradigm of things of the form $x$ is $F$ are propositions. Now one might try positing worldly things that are somehow of the form $x$ is $F$ but which are not propositions. For example, Fish develops his naive realism in a way that has states of affairs (or facts) among the relata of the fundamental perceptual relation (2009: 51-54). Indeed, one might be tempted to think that facts, or states of affairs, are not propositions but are rather the things to which propositions correspond or which are the “truthmakers” of propositions. Here naive realists are very much hostage to metaphysical fortune: the standard view is that facts are true propositions. After all, the standard view is that knowledge is of facts and that the knowledge relation is to the same kind of thing as belief. Since belief is a relation to propositions the standard way of putting all this together is to think of facts as true propositions. This is another reason the naive realists should adopt tropes as the entities of awareness. Tropes like propositions bind objects and properties together but unlike facts, it is standardly thought that tropes are not propositions.

You might think that there is still a kind of binding problem for the naive realist who adopts tropes.\textsuperscript{57} Suppose a fence, Sally, is white. There is a trope, Jimmy, that is the instantiation of whiteness in Sally. You see Sally. You see Jimmy. What guarantees that

\textsuperscript{56} See the bridge principle for naive realism for analyzing facts of the form, $s$ perceives $x$ to be $F$, in chapter 2 as an illustration.

\textsuperscript{57} Thanks to Adam Sennet who raised this worry.
you now see Sally as white? Couldn’t you misperceive Jimmy as belonging to another fence, Charlie?

It may be that the best way forward for the naive realist is to hold that you perceive Sally by perceiving one of its tropes.

Object Perception Principle: You count as perceiving an ordinary object in virtue of standing in the fundamental perceptual relation to a trope of which that object is a participant.

That principle doesn’t violate directness since, after all, tropes are mind-independent constituents of reality and are plausibly concrete since they are down here in the world rather than out there in Platonic Heaven. Note that events are plausibly concrete and many, including Jonathan Bennett (1988, 2002) and Campbell (1981), claim that events are tropes. But it does remove some of the intuitive appeal of naive realism since their account is often motivated by the idea that we stand in fundamental perceptual relations to cats, dogs, people, and chairs, etc.  

This clarification regarding perceiving tropes in veridical perception will be increasingly important as we move through the rest of this chapter and the next. Further, while I have made a case that naive realism works best with tropes in the process of considering edenic realism, the argument for tropes generalizes to any form of naive realism that seeks an account of phenomenal character based on standing in perceptual

---

58 I am being concessionary here since tropes are sometimes described as “abstract particulars” which does not suggest they are concrete.

59 There is another issue. Why should seeing Sally guarantee the fence looks white. Even if Sally is an instance of whiteness couldn’t there be the following kind of illusion: A subject, S, has a trope illusion when S stands in the fundamental perceptual relation to a trope that is in fact an instance of F-ness but one sees the bearer of F-ness as G, where being G precludes being F. I leave this as a challenge to the naive realist who adopts tropes. This is a variant on the problem of error discussed in chapter 3.
awareness relations to particular mind-independent entities. Next, I return to the putative counterexamples to slogan M.

1.5.5 Accounting for Variation and Error

The edenic realist, now armed with tropes (in addition to objects) as standard entities of awareness in veridical perception, can hold that ordinary objects have many mind-independent tropes. Take the case of two subjects disagreeing over whether Munsell chip G is unique green. The edenic realist might argue that G has many color tropes. It has the unique green trope, T1 and it has the greenish-yellowish trope, T2. Both subjects stand in the relation of perceptual awareness to G. But only one subject stands in the relation of perceptual awareness to T1 and the other stands in the relation of perceptual awareness to T2. Both count as veridical perceptions. So far, so good.

But what happens when we consider degenerate cases that involve perceptual error? For example, misperceiving a rope as a snake or hallucinating a pink rat where there is none. If the edenic realist posits tropes in the world to account for character in non-veridical experience, it looks like we have a new problem in accounting for perceptual error since one can argue that all perceptual experience is veridical. There is a trope that determines the character somewhere or other: either in an object, a sense datum, a region of space, or somewhere in the brain. Perhaps there was no perceptual error in Eden. But we are no longer in Eden.

This by no means is a death blow to edenic realism. But it gets dicey enough that we might look for other routes forward. I’ll return to this dilemma for tropes and naive
realism in chapter 3.\textsuperscript{60} For current purposes, let’s continue explicating the phenomenological thesis recognizing that edenic realism will need to overcome various problems associated with perceptual variation.

1.5.6 Standpoint

Contemporary naive realists are not defending a form of edenic realism. They are more sophisticated. They regularly incorporate both a point of view and environmental conditions into their story about character.\textsuperscript{61} But incorporating additional (external) conditions about the subject’s position and surroundings will not meet all the related challenges. There is growing empirical evidence that character varies not just with the subject’s point of view, environmental conditions, and the entities perceived, but also on the basis of internal psychological and neurological conditions of the subject: their conceptual resources, facts about their post-retinal neuro-computational processing, idiosyncrasies of their sensory anatomy and sensory systems, and the distribution of their attention resources.\textsuperscript{62} The challenge for contemporary naive realists is how to accommodate these findings (and the putative counterexamples to edenic realism).\textsuperscript{63} While this is an important topic of emerging interest in the philosophy of perception today, to go

\begin{itemize}
\itemsep0pt
\item \textsuperscript{60} One horn of the dilemma is the need for tropes in an account of character, objects alone are not enough. The other horn of the dilemma is once you admit tropes into your story of veridical character, how do we explain character in non-veridical cases?
\item \textsuperscript{61} There is also a trend to also integrate variations in the relation of perceptual awareness in the account to accommodate a relation of degraded perceptual awareness. See Brewer (2011: 116), French (2014), and Campbell and Cassam (2014: 28). I’m not inclined to go that way. My preference would be to keep the perceiving relation the same and instead use another variable in the standpoint set of variables. See below.
\item \textsuperscript{62} For arguments that the phenomenological thesis is incompatible with empirical research, see Pautz (2011, 2014, 2017, Forthcoming). For discussion and a naive realist response to these issues, see Beck (2019).
\item \textsuperscript{63} For examples of naive realists attempting to accommodate internal conditions of the subject: see Brewer (2011: 121-122), Campbell (2002: 118-119), Fish (2009: 75).
\end{itemize}
further into the details of this debate will take us too far afield. Instead, we will accept that some (or all) of the relevant factors mentioned will need to be incorporated in a complete naive realist story of how character is determined. Thus, my goal is to present a way of understanding the phenomenological thesis suitably weak enough for all naive realists to agree as a starting point, accepting that some naive realists may go further in advancing a stronger version.

Given the discussion above, I’ll use the term ‘standpoint’ as a catch all to include the subject’s place, point of view, relevant internal psychological and neurological properties, and the relevant environmental conditions and physical relations to the entities of awareness. Next, consider the following slogan:

\[(RM) \text{ Given the same standpoint, a match in entities of awareness, guarantees a match in character.}\]

In attempting to seek a weak reading of the phenomenological thesis by including various additional constraints, there is a worry about whether mind-independent entities will be playing any role in determining character at all. If standpoint alone were sufficient to determine character, we would get the following slogan:

\[(SM) \text{ A match in standpoint, guarantees a match in character.}\]

I take it SM is inconsistent with the underlying motivations for naive realism. We might invoke the following constraint on the appropriate informational “thickness” of a

---

64 A standpoint is a set with the values of each of these additional measures. Campbell (2009) and Beck (2019) also use ‘standpoint’ for a similar purpose.
standpoint: A standpoint should not include so much information that it makes SM true, otherwise it renders RM trivial.

While RM is schematic at best, it serves as a heuristic in deepening our understanding of the phenomenological thesis to some extent. Be that as it may, there is still the second question concerning what is meant by ‘determines’ in the phenomenological thesis. I turn to that next.

1.5.7 Modal Corollaries

Campbell uses the word ‘bound’ and Martin uses ‘determines’ to express the relationship between the character of a subject’s experience and the entities of awareness (plus the subject’s standpoint). I’ve followed Martin in using ‘determines’ in the phenomenological thesis and used the term ‘guarantees’ in the slogans. We can further our understanding of the phenomenological thesis by rendering ‘determines’ (and ‘guarantees’) in such a way that it has modal implications. I’ll proceed by example. Given the forgoing discussion, I propose the following corollary of the phenomenological thesis:

(PTC 1) Necessarily, if a subject undergoes a veridical experience from standpoint, $R$, and is perceptually aware of all and only entities, $x_1 \ldots x_n$, and has an experience with character $C$, then necessarily, any subject who is undergoing a veridical experience from standpoint, $R$, and is perceptually aware of all and only $x_1 \ldots x_n$ will have an experience with character $C$.

PTC 1 addresses the question of sameness in character. Given all the potential conditions in play, it potentially sets a high bar for a single subject to have two experiences with the
same character, and an even higher bar for two different subjects to each have an experience with the same character. To be clear, my goal here is to further our understanding of the phenomenological thesis by showing the modal implications. PTC 1 may help. But what about differences in character?

Martin, in the last full sentence of the passage quoted above, is illustrating what he means by ‘determination’ with what I take to be a modal implication concerning difference in character. First, consider the slogan:

(MM) A mismatch in entities of awareness, guarantees a mismatch in character.

Next, a modal version, that also includes standpoints:

(PTC 2) Necessarily, if a subject undergoes a veridical experience from standpoint, $R$, and is perceptually aware of all and only entities, $x_1…x_n$, and has an experience with character $C$, then necessarily, any subject who is undergoing a veridical experience from standpoint, $R$, and is not perceptually aware of all and only $x_1…x_n$ will not have an experience with character $C$.

This second corollary of the phenomenological thesis captures Martin’s consequence of taking ‘determination’ talk literally. One counterintuitive result is that two experiences of “qualitatively matching scenes” with numerically distinct objects will necessarily have different characters. Let’s take a toy example of looking at a black ball in front of a white screen. Further, there are two black balls; they are qualitatively identical but numerically distinct. Next, suppose the standpoint remains fixed but we swap out the balls over the
course of two experiences for the same subject. Intuitively, the two experiences yield the same character, but PTC 2 entails the character of the two experiences is different.

I want to make a further point about this sort of counterintuitive result. The externality thesis states that concrete objects are constituents of the events that are the experience. If we assume that experiences with different constituents are distinct experiences, then any difference in concrete objects of awareness yields a different experience. But experiences are particulars. We don’t type experiences by their concrete constituents. We type experiences by their character. So, it doesn’t follow from externality and the phenomenological thesis alone that two experiences with distinct concrete constituents must differ in character. Rather, that result follows from rendering the phenomenological thesis as entailing PTC 2 and also counting concrete objects as some of the entities, $x_1 \ldots x_n$.

The counterintuitive result can be avoided. I’m going to treat talk of ‘qualitatively identical objects’ as referring to objects that share all and only the same properties. But since it is not the case that the instantiation of any property always potentially contributes to character, one has to be careful. Presumably, some properties are perceptually inert, in that they make no (possible) contribution to character. So, in this section when I use property (and trope) talk unqualified, I will mean all properties minus the perceptually inert properties.

We are in search of a corollary of the phenomenological thesis to replace PTC 2 and accommodate our intuitions concerning qualitative similarity. In order to present the next corollary, I will first introduce a few notions using trope-theory. The first notion is that of one entity being a partner of another entity. I’ll introduce the notion of $x$ being a
partner of $y$, through cases. Every object is a partner of every other object. No qualitative property or relation is a partner of anything except itself. Every trope that consists in $x$ instantiating $F$ (where $F$ is a qualitative property) has as a partner any trope that consists in some $y$ instantiating $F$. And so on.\(^{65}\) Now, we can say that where $x_1 \ldots x_n$ are the entities of perceptual awareness of $S_1$ and $y_1 \ldots y_n$ are the entities of perceptual awareness of $S_2$, a one-to-one mapping of $x_1 \ldots x_n$ to $y_1 \ldots y_n$ is a kosher permutation just in case every entity gets mapped to a partner of it. With the notion of a kosher permutation in place, here is the third corollary of the phenomenological thesis:

(PTC 3) Necessarily, if a subject undergoes a veridical experience from standpoint, $R$, and is perceptually aware of all and only entities, $x_1 \ldots x_n$, and has an experience with character $C$, then necessarily, any subject who is undergoing a veridical experience from standpoint, $R$, and is not perceptually aware of a kosher permutation of $x_1 \ldots x_n$, will not have an experience with character $C$.

Now, if we go back to the toy case of looking at a black ball on a white screen, we avoid the counterintuitive result. Since the experiences involve kosher permutations, it doesn’t entail the two experiences necessarily differ in character. But they don’t necessarily have the same character either. For that result, we would need the following:

(PTC 4) Necessarily, if a subject undergoes a veridical experience from standpoint, $R$, and is perceptually aware of all and only entities, $x_1 \ldots x_n$, and has an experience with character $C$, then necessarily, any subject who is

\(^{65}\) I haven’t covered all the cases, but the idea should be clear enough.
undergoing a veridical experience from standpoint, R, and is perceptually aware of a kosher permutation of $x_1 \ldots x_n$, will have an experience with character $C$.

Note that none of these corollaries tells us how exactly the entities of awareness nor the standpoint contribute to character (Beck 2019: 610). Throughout the dissertation, I’ll assume the phenomenological thesis implies PTC 3 and PTC 4.

The last point I want to make in this section is that the phenomenological thesis neither entails nor is entailed by the ontological thesis. The two theses are logically distinct. But they complement one another. First, the phenomenological thesis provides the naive realist answer to the character question. Second, the character of a veridical perception is determined by the same mind-independent entities that feature as relata in the perceptual awareness relation in the ontological thesis. Further, from this foundation, it is a natural step to argue that one reason the relation of perceptual awareness is fundamental is that the entities standing in the perceptual awareness relation partly determine character. In the next chapter, I’ll explore the primary motivation for adopting naive realism.

1.6 Concluding Remarks

After an introduction to this dissertation (1.1), I covered two preliminary matters: the categories of perceptual experience (1.2) and phenomenal properties (1.3). Next, I presented the core theses of naive realism (1.4) and in so doing also provided an account of fundamentality in order to clarify the special status of the relation of perceptual awareness. In the last section (1.5), I explicated the phenomenological thesis and argued for a modal interpretation of the determination relation. An important result in this chapter
is that naive realism is likely best off adopting an ontology that includes tropes, over and above mind-independent (ordinary) objects, in order to maintain their distinctive position on how character is determined in veridical perception.
Chapter 2
Intuitive Appeal

In this chapter, I explore the intuitive appeal of naive realism. Arguably, this is the primary motivation for the view. By ‘intuitive appeal’ I mean this: that pre-theoretical reflection provides prima facie support for a view. So, the intuitive appeal of naive realism is that pre-theoretically reflecting on perceptual experience supports naive realism.

I begin this chapter by covering various preliminary matters, including a discussion on the approach I take and a summary of findings from pre-theoretical reflection—which I call presentationism (2.1). Next, I go on to evaluate whether presentationism supports the theses of naive realism over rival theories, beginning with the directness thesis of perception (2.2). This is followed by a brief discussion of the ontological and externality theses (2.3), and then a lengthier discussion of the phenomenological thesis (2.4). Last, I summarize conclusions from the preceding sections (2.5).

2.1 Preliminaries

In recent years, several novel arguments have been made for naive realism. Martin (2002) argues that the phenomenology of sensory imagination supports naive realism over rival views. Campbell (2002, 2011) and Brewer (2004, 2006, 2011) both argue in their own way that the naive realist account of the character of experience provides the best explanation of our possession of concepts for the objects and properties we perceive. Allen (2020) presents a transcendental argument for naive realism. These arguments are complex
and controversial. It will be more helpful to focus on what many take to be the primary motivation for defending naive realism, that is, its intuitive appeal.

2.1.1 Pre-Theoretical Reflection

Several attempts have been made to delineate the reasoning behind the intuitive appeal of naive realism.\(^6\)\(^6\) One approach would be to attempt to translate those arguments into a uniform terminology and then go on to compare, categorize, and evaluate those arguments. In what follows, I take a simpler approach. First, I’ll ask:

What do we find by pre-theoretically reflecting on perceptual experience?

In providing an answer to the question, I’ll attempt to arrive at findings through uncontroversial data points. Second, I’ll investigate what those findings show.

Before reviewing findings from pre-theoretical reflection, a caveat about the modality of experience. It may be that there are important relevant differences in our reflective findings between the senses. Be that as it may, the focus here will be on visual experience. We can take our findings to either extend to total multi-modal perceptual experience (which includes visual phenomenology) or to only support theories of visual experience. Now, without further ado, let’s reflect on the phenomenology of visual experience.

When we open our eyes to the scene before us and have a visual experience, it is as if the shutters of a window were opened and ordinary objects (e.g., trees, dogs, oranges, oranges...)

tables, etc.) in the environment are presented. Call this the window model. There is a phenomenological presence to the objects perceived in visual experience that is lacking when we are having a thought about the same object or when we are imagining the object. Further, the phenomenology of seeing ordinary objects in the environment strikes us as direct: that such visual experiences (of ordinary objects) is not occurring in virtue of (or through) seeing a mental image. In contrast to the window model of perception, there is the picture model. According to the picture model, we see ordinary objects in the environment in virtue of (or through) seeing something else, like a mental image. When we look around and reflect on our visual experience, our experience supports the window model over the picture model.

2.1.2 Presentationism

In sum, when we reflect on the phenomenology of visual experience, the following seems to be true: (1) our experience is of ordinary objects, (2) some of the objects of experience have a distinctive quality of presence which is lacking when having thoughts, beliefs and episodes of imagination about those same objects, and (3) perceiving ordinary objects is direct, in the sense that we are not aware of ordinary objects in the environment by being aware of something not in our environment. I’ll call the conjunction of these three findings, presentationism.\(^{67}\)

\(^{67}\)To be clear, presentationism is distinct from the controversial transparency thesis which states (i) that phenomenal properties are properties of the objects perceived (or properties of the worldly propositions about the objects perceived), and (ii) that we are never aware of purely experiential properties (mental paint)-properties that are not properties of objects nor properties of worldly contents. One might argue that presentationism supports the transparency thesis. I’m not going to get into the details of the debate over transparency and various versions of the thesis. The main problem for a standard version of the transparency thesis is that reflecting on our experience yields counterexamples like the following: I sometimes experience a visual disturbance prior to the onset of a migraine. In such cases, my visual experience is such that there
2.1.3 On Method

Next, I’m going to assume that when we pre-theoretically reflect on experience, something like presentationism seems to be true. That raises two further questions:

(1) If presentationism is true, what does that show about naive realism?

(2) Granting that presentationism seems to be true upon pre-theoretically reflecting on experience, how much evidence is that for presentationism.

Of these two questions, I will focus on the first. While the second question is intriguing, I will not say much about it here (other than make an important clarification regarding the nature of pre-theoretical reflection, which I will do next).

As I have presented it above, pre-theoretical reflection involves either introspecting on the phenomenology of one’s own perceptual experiences or about one’s beliefs thereof. As such pre-theoretical reflection is distinct from the sort of scientific and philosophical reflection engaged in when theorizing about perceptual experience. Scientific and philosophical reflection involves many more considerations than the deliverances of pre-theoretical reflection. Obviously, perceptual experience may not actually be the way it seems to be before doing science and philosophy. Thus, presentationism may or may not survive scientific and philosophical reflection (Hawthorne and Kovakovich 2006: 180).

are streaks of distorted colors that blur the periphery of my visual field around the focal point of attention creating a peculiar visual aura. And yet it has never seemed to me as if the features of the visual disturbance are properties of mind-independent concrete objects in my environment. The color streaks always strike me as “internal features” of my experience rather than that they are happening “out there” in the mind-independent world. These experiences immediately lead to the belief that something abnormal is going on with me internally, with my perceptual system and brain, rather than something abnormal happening in my environment. This type of visual disturbance associated with migraines is not uncommon and well documented (Hupp, Kline and Corbett 1989). A similar observation can be made by reflecting on more pedestrian cases of blurred vision (Smith 2008). There is a voluminous literature on the topic of transparency, the most influential include Tye (2009, 2015), Block (1996, 2010), Harmon (1990), and Speaks (2009).
Presentationism (or any view in the vicinity) is often labeled a “common sense” or “vulgar” view because it is pre-theoretical. Now, it may be that phenomenological experts (e.g., philosophers of perception) upon reflecting on their own experience validate the vulgar view by converging on presentationism. But that doesn’t make the support for presentationism any less *prima facie* or pre-theoretical in the sense under discussion. It is a virtue of a complete theory of perception that it accommodates the pre-theoretical allure of presentationism --- i.e., explain why presentationism seems to be true to pre-theoretical reflection. But it is a further step to suppose that a complete theory will accommodate the allure of presentationism by accepting the truth of that thesis.

2.1.4 Barebone Intentionalism

In order to evaluate whether presentationism supports naive realism, I’ll take each thesis of naive realism in turn. Further, if it turns out that presentationism supports some but not all of the four theses of naive realism, it is then worth asking how the other main theories of perception fare in light of the findings. I will be suggesting that a version of the view known as *intentionalism* is left unscathed by pre-theoretical reflection, so it is worth saying a little at the outset about the barebones of that approach.

*Intentionalism* is the dominant theory of perception today. Accordingly, it is the main rival to naive realism. There are many varieties of intentionalism, but all intentionalists about perception hold that the nature of perceptual experience is for a subject

---

68 Bourget and Chalmers (2014), question 21. I will use ‘intentionalism’ throughout this dissertation instead of the more traditional ‘representationalism’. Intentionalism is also called ‘the content view’. For my purposes, these are mere terminological differences. The primary focus of the dissertation is an internal debate between naive realists over the nature of hallucination. The comparison with intentionalism is mostly for heuristic purposes. Prominent intentionalists include Byrne (2001), Dretske (1995), Harman (1990), Pautz (2010), Seigel (2010), and Tye (1992).
to stand in a relation, call it *perceptually entertaining*, to a proposition. According to intentionalism, perceptual entertaining (a relation to propositions) is a more fundamental relation than perceptual awareness (a relation to objects). This is in conflict with the ontological thesis. Second, intentionalists hold that the character of a perceptual experience is determined by its content. In a slogan, there is no difference in character without a difference in content. As with naive realism, one might instead try out the thesis that differences in content guarantee differences in character but not the other way around. But as a matter of fact that variant is not the prominent view. When I use the term, ‘intentionalism’, unqualified, I mean to pick out any version of the view with the aforementioned commitments.

2.2 Presentationism and Directness

Does presentationism support the directness thesis? For convenience, here is the directness thesis again:

When a subject stands in the relation of *perceptual awareness* to mind-independent concrete entities, that is not (even in part) in virtue of standing in the perceptual awareness relation to entities that are not mind-independent concrete entities, e.g., sense-data, mental images, propositions, etc.

First, we might ask whether or to what extent the technical terms in the directness thesis align with the commonsense terminology used in the characterization of presentationism. I’m going to make a few simplifying assumptions throughout this comparison. First, I’ll

---

69 The name ‘perceptually entertaining’ is inspired by Pautz’ (2007, 2010) ‘sensorily entertaining’.
assume that ordinary objects are mind-independent concrete entities. Second, that any entity that is not a mind-independent concrete entity is either a mental entity (mental image, mental representation, sense-data, etc.) or a pure abstract entity (proposition, universal, etc.), and as such they are not in the concrete environment of perceiving subjects.

Next, if we also assume that ‘aware of’ as used by the vulgar picks out the relation of perceptual awareness, then we can say that the directness thesis plus the simplifying assumptions entail the following corollary:

(DTC 1) When a subject is perceptually aware of ordinary objects in the environment, that is not in virtue of being perceptually aware of something not in the subject’s environment.

And this is nearly identical to the description of direct perception in presentationism. Next, consider another corollary of the directness thesis (and simplifying assumptions):

(DTC 2) When a subject is perceptually aware of unordinary objects in the environment, that is not in virtue of being perceptually aware of something not in the subject’s environment.

The term ‘unordinary objects’ is meant to include anything that counts as a mind-independent concrete entity that is not an ordinary object. For an example of an unordinary object, consider the subject-facing perceivable surface of an ordinary object as a plausible candidate. Since presentationism is a view about ordinary objects, presentationism neither supports nor is supported by DTC 2. But it nevertheless is compatible with it. Note, given
the simplifying assumptions, DTC 1 and DTC 2 together entail the directness thesis, and
the directness thesis entails DTC 1 and DTC 2.

Now, we can clearly see how presentationism supports the directness thesis. Presentationism affirms DTC 1 which is entailed by the directness thesis. Further, there is nothing about presentationism that is incompatible with DTC 2. Accordingly, we might say presentationism provides partial support for the directness thesis. Now, if a naive realist denies the existence of unordinary objects or denies that we are ever perceptually aware of them, then the truth of presentationism (together with the simplifying assumptions) guarantees the truth of the directness thesis. Next, I want to contrast this result with rival theories of perception.

2.2.1 Directness and Rival Theories

The sense-datum theory of perception rejects the directness thesis. According to the sense-datum theorist, we perceive ordinary objects in our environment in virtue of being directly aware of sense-data. Prominent sense-datum theorists, such as Jackson (1977) and Robinson (1994) treat sense-data as mental entities. Since sense-data, understood as mental entities, are not in the physical environment of the perceiving subject, presentationism is incompatible with the sense-datum theory. Thus, in considering the

\[\text{\footnotesize{\textit{Footnote:} In the text, I’m ignoring a variant of the view that treats a sense-datum as a peculiar mind-independent concrete object in the environment, considered in Moore (1919) and Bermúdez (2000). Such a variant may at first appear compatible with the directness thesis, as I have formulated it. But it is not. On this account we only stand in the most fundamental perceiving relation to unordinary objects (the surface of objects) and not also to ordinary three-dimensional objects, contrary to presentationism and naive realism. So, it rejects DPC 1 and accepts only DPC 2.}}\]
direct perception of ordinary objects, presentationism supports naive realism over the sense-datum theory.

Next, I’ll consider whether the directness thesis is compatible with intentionalism. Since intentionalists take perceptually entertaining a proposition as the most fundamental perceptual relation, it is tempting to assume that intentionalists also hold that we stand in the perceptual awareness relation to ordinary objects in the environment (at least partly) in virtue of standing in the perceptual entertaining relation to a proposition. This view may not strictly follow from the fundamentality claim alone. Nevertheless, let’s go along with this plausible understanding of intentionalism for now and assume the following is a tenet of intentionalism:

(ITP) Subjects stand in the perceptual awareness relation to ordinary objects in the environment (at least in part) in virtue of standing in the perceptual entertaining relation to a proposition.

Next, I will show how ITP is compatible with presentationism and the directness thesis.

Showing how presentationism is compatible with intentionalism is important since it is often said they are incompatible views. Here is an example from Crane and French:

Thus, like sense-datum theorists, ...intentionalists reject Direct Realist Presentation [that perceptual experiences are direct perceptual presentations of ordinary objects], and admit that we are not ever directly presented with ordinary objects, not even in veridical experience. (2021: section 3.3.4)

To be clear, what Crane and French call ‘Direct Realist Presentation’ is a version of presentationism based on pre-theoretical reflection since they claim it is “embedded within our ordinary conception of perceptual experience” (2020: section 1). It is worth also
getting clear on what it is for a subject to be ‘directly presented with ordinary objects’. I contend ‘directly presented with’ is just the relation of perceptual awareness, since Crane and French describe it as involving “a special intimate perceptual relation to an object” (2020: section 1.3). Next, I’ll attempt to shed some light on this common point of confusion.

First, notice there are at least two different perceptual relations at play here: (a) \textit{perceptual awareness}, a relation that holds between subjects and ordinary objects in the environment (possibly among other types of entities such as events and tropes), and (b) \textit{perceptual entertaining}, a relation that holds between subjects and propositions. Second, the directness thesis of naive realism is solely concerned with the relation of perceptual awareness and is entirely quiet on the relation of perceptual entertaining. Thus, the directness thesis and ITP are not in conflict and so are compatible.

To be fair, there may be a simple form of intentionalism that holds that perceptual awareness and perceptual entertaining are not two distinct relations but in fact the same relation. I have reservations about a single perceptual relation taking such ontologically different entities as \textit{relata}, but if that were so, then the directness thesis would be incompatible with simple intentionalism. Further, simple intentionalism would be in tension with presentationism since propositions are presumably not something in the subject’s environment.

But presentationism, so far at least, is compatible with the sophisticated version of intentionalism that makes a distinction between the two relations. After all, an intentionalist can affirm their commitment to DTC 1 simply in virtue of holding that we don’t stand in the same perceptual relation to ordinary objects as we do to propositions.
Further, even if the intentionalist adheres to ITP, it is not clear that pre-theoretical reflection has anything at all to say about a claim as theoretical as: *partly in virtue of sensorily entertaining a proposition*. We shouldn’t expect presentationism, as a view which is derived from and supported by pre-theoretical reflection, to provide evidence for or against theory-laden claims that include technical concepts.

### 2.3 Vulgar Quietism on Fundamentality and Externality

Considering the limits of pre-theoretical reflection, it is reasonable to claim that presentationism has nothing to say about the ontological thesis as it concerns the metaphysics of fundamentality and the entire theoretical domain of perception. The verdict is the same for the relevant ontological analog theses made by the sense-datum theorist and the intentionalist. Similarly, presentationism is silent on the externality thesis as it concerns the metaphysics of experiences, events, and constituency. Next, I’ll turn our attention to the phenomenological thesis.

### 2.4 Pre-Theoretical Reflection and Phenomenology

Since presentationism is based on pre-theoretical reflection on the phenomenology of visual experience, we might expect presentationism to strongly support the phenomenological thesis of naive realism. For convenience, here is the phenomenological thesis again:
In veridical perception, the character of experience is partly determined by the particular mind-independent entities to which the subject stands in *perceptual awareness*.

After considering several modal interpretations of the phenomenological thesis (section 1.5.7), I argued for a rendering of ‘determines’ that led to accepting PTC 3 and PTC 4 as corollaries implied by the phenomenological thesis. But since both corollaries employ concepts that are theoretically sophisticated, it will be difficult to find support from pre-theoretical reflection for those corollaries. Instead, let’s start with presentationism and see whether it provides any support for a version of the phenomenological thesis. Again, for convenience, here are the tenets of presentationism:

1. our experience is of ordinary objects,
2. some of the objects of experience have a distinctive quality of presence which is lacking when having thoughts, beliefs and episodes of imagination about those same objects, and
3. perceiving ordinary objects is direct, in the sense that we are not aware of ordinary objects in the environment by being aware of something not in our environment.

Taking each claim above at face value, there is nothing explicitly related to character. I constructed presentationism on the basis of pre-theoretically reflecting on what it is like as a subject to undergo visual experiences. But I did not overtly include a tenet about what determines character. I’ll address that next.
2.4.1 Vulgar Phenomenology

We can return to pre-theoretically reflecting on veridical visual experience, as we did before, and explore the nature of phenomenal character. It is helpful here to compare experiences. What do we notice about what a visual experience is like when seeing an orange versus what an experience is like when seeing a banana? There are differences in shape and color between the orange and banana which make a difference in what each visual experience is like subjectively. When we pre-theoretically reflect on our veridical visual experience and base our inquiry on the following question:

What determines what this visual experience is like?

A simple answer to the question on behalf of the vulgar is:

What we perceive objects to be like determines what our experience is like.

If we set aside worries about the vagueness of the answer, it is uncontroversial that this is the view of the vulgar. I’ll assume it is acceptable to replace the expression ‘what an experience is like’ with the expression ‘the character of experience’. Given these assumptions, pre-theoretical reflection provides support for:

(Proto-PT) In veridical perception, the character of experience is determined by what the subject perceives objects to be like.

Next, I’ll show how (Proto-PT) provides support for the phenomenological thesis.
2.4.2 Bridging the Vulgar Divide

There is a divide between the vulgar view on character, understood as (Proto-PT), and the phenomenological thesis of naive realism (NR-PT). The difference between the two theses is in what they each name as doing the partial determining. For comparison, here is what follows ‘determined by’ in each thesis:

(NR-PT) the particular mind-independent entities to which the subject stands in perceptual awareness,

versus

(Proto-PT) what the subject perceives objects to be like.

In order to bridge the divide, I will introduce more theory since it is unlikely that more pre-theoretical findings alone will show how support for (Proto-PT) is support for (NR-PT). Let’s begin by analyzing the expression ‘what a subject perceives objects to be like’. Given the qualitative nature of ‘to be like’, it is reasonable to assume that what a subject perceives object to be like is determined by facts of the form:

A subject, s, perceives object, o, to have property F.

A subject, s, perceives object, o₁...oₙ to be standing in n-place relation R.

‘Perceives’ as used here implies perceptual success. Thus, in the veridical case, the properties of objects we perceive (and relations between objects we perceive) determine what we perceive objects to be like. This is the first bit of theory required to bridge the divide.
Next, I’m going to introduce bridge principles. The naive realist has a preferred analysis of facts of the form just presented above. These are naive realist bridge principles:

\[(\forall S)(\forall o)(\forall F) ((\text{a subject } S \text{ perceives an object } o \text{ to have property } F) \text{ if and only if } (\exists z)((z \text{ is an instance of the property } F \text{ in } o ) \land (S \text{ is perceptually aware of } z))\]

and for a two-place relation,

\[(\forall S)(\forall x)(\forall y)(\forall R) ((\text{a subject } S \text{ perceives an object } x \text{ to stand in relation } R \text{ to } y) \text{ if and only if } (\exists z)((z \text{ is an instance of the relation } xRy ) \land (S \text{ is perceptually aware of } z)) \land (S \text{ is perceptually aware of } x)),\]

and so on.

The order of ‘x’ and ‘y’ in ‘xRy’ matters. Take the case of Ted kicking Dean. An instance of Ted kicking Dean is not the same as an instance of Dean kicking Ted. The bridge principle for three and greater place relations follow the same pattern.

Since an instance of a property is a trope, z is a trope. Therefore, what a subject, s, perceives an object to be like is determined by the tropes to which s stands in perceptual awareness. Putting all this together, yields:

\[(\text{Trope-PT}) \text{ In veridical perception, the character of experience is determined by which objects and tropes the subject stands in perceptual awareness.}\]

Notice, Trope-PT is very close to NR-PT. There are two differences. First, Trope-PT omits ‘partly’ in specifying the determination relation. In chapter 1, I explained that ‘partly’ is used in the NR-PT to accommodate various considerations I lumped into a standpoint. One could make the argument that the vulgar, without philosophical or
scientific training, appreciate the fact that the condition of one’s eyes contribute to character.

The second difference is in the entities named: the very general ‘mind-independent entities’ in NR-PT versus specific types of entities in Trope-PT. Since objects and tropes are both particular mind-independent entities, Trope-PT is a specific version of the phenomenological thesis. On a version of naive realism where those are the only two types of entities to which we ever stand in relations of perceptual awareness, then the Trope-PT and the phenomenological thesis would be equivalent. Thus, pre-theoretical reflection supports the phenomenological thesis of naive realism. Is this a unique advantage of naive realism? I turn my attention to this question next.

2.4.3 Intentional Bridging

Does intentionalism similarly bridge the divide between (Proto-PT) and their thesis about what determines character? First, here is a generic version of an intentionalist phenomenological thesis:

(I-PT) In veridical perception, the character of experience is determined by which proposition the subject is perceptually entertaining.\(^{71}\)

Next, in the same manner as the naive realist above, we can wheel in bridge principles on behalf of the intentionalist:

\(^{71}\) Intentionalism can use this same thesis generally for all perceptual experience. I’m doing it this way to preserve the symmetry in comparison.
\[(\forall S)(\forall x)(\forall F)\ [(\text{a subject } s \text{ perceives an object } o \text{ to have property } F) \text{ if and only if } S \text{ perceptually entertains that } Fo,\]

and for a two place relation,

\[(\forall S)(\forall x)(\forall y)(\forall R)\ [(\text{a subject } S \text{ perceives an object } x \text{ to stand in relation } R \text{ to } y) \text{ if and only if } S \text{ perceptually entertains that } xRy,\]

and so on.

Given the analysis of ‘what a subject perceives an object to be like’ above, and the intentionalist bridge principles, the intentionalist can show an equivalence between Proto-PT and I-PT. Thus, if pre-theoretical reflection supports Proto-PT, the intentionalist can show how it also supports I-PT. Given the greater simplicity of the intentionalist bridge principles over the naive realist principles, the argument that pre-theoretic reflection supports intentionalism is less contorted.

2.5 Concluding Remarks

In this chapter, I have attempted to provide a clear interpretation of the intuitive appeal of naive realism. I have shown that pre-theoretical reflection supports naive realism. Specifically, pre-theoretical reflection supports the directness thesis and the phenomenological thesis. Unsurprisingly, pre-theoretical reflection supports naive realism over the sense-datum theory. Perhaps surprisingly, pre-theoretical reflection provides equal support for intentionalism. Given there is nothing naive about tropes, one might argue that pre-theoretical reflection actually favors intentionalism.
Chapter 3
Hallucination and the Screening Off Problem

There are many different arguments from hallucination for and against different views in the philosophy of perception. These different arguments are all concerned with how best to account for hallucination in our philosophical theory of perception. If we think we should provide a similar account for the three main categories of perceptual experience (veridical perception, illusion, and hallucination), then cases of hallucination are a challenge for naive realism. Since naive realists wish to account for phenomenal experience by appealing to relations to mind-independent entities in the environment, the challenge is that hallucinations seem to have character but such relations to mind-independent entities are absent. The core theses of naive realism, as presented in chapter 1, are restricted to cases of veridical perception. So, the naive realist still owes us an account of hallucination if they seek to provide an adequate philosophical theory of perception. Further, the naive realist owes us an account of the character of hallucinatory experience compatible with the seamless transition case. In my view, answering the character question for hallucination is the crux of the various challenges and arguments from hallucination for naive realism.\textsuperscript{72} Accordingly, the character question will be the focus of this chapter.

I begin this chapter taking up preliminary matters which mostly involves providing a taxonomy of views, custom tailored for the discussion that follows (3.1). Second, I

\textsuperscript{72} One might object that there are arguments from hallucination that are focused on the ontological nature of experiences or the entities of awareness in hallucination, rather than character per se. But even those arguments, generally turn on a premise concerning the sameness of experience typed by character or by introspective indistinguishability of character. This is one of several reasons why I focus on the character question and the arguments that follow in this chapter.
present a challenge for a straightforward approach to the character question, which I call, ‘common factor naive realism’ (3.2). Next, I present both Martin’s original screening off problem and a new version based on narrowness against naive realism (3.3). Finally, I present and evaluate replies to the screening off problem (3.4) and end the chapter with concluding remarks (3.5).

3.1 Preliminaries

Before jumping into the character question for hallucination, I want to make some important distinctions that divide approaches to theorizing about the character of perceptual experience. First, I’ll distinguish between relationalist and non-relationalist approaches to the character of hallucination (3.1.1). Next, I’ll distinguish between common-factor views and disjunctivist views (3.1.2). Last, I’ll distinguish between positive and negative forms of disjunctivism (3.1.3).

3.1.1 Relationalism

There is a basic distinction between what we might call relationalist and non-relationalist approaches to the character of perceptual experience. The relationalist wishes to somehow explain the character of perceptual experience in terms of perceptual relations to entities (whether they be entities in the mind or entities in the world, whether they be mundane concreta, or sense data, or tropes or propositions).73 The non-relationalist, paradigmatically

---

73 To be clear, I’m using ‘relationalist’ in a different way than it is often used. It is often used to draw a distinction between naive realism and its main rivals. For example, Campbell (2002) calls his naive realism, ‘the relational view’. Compare Pautz (2010: 259-260) who makes a similar distinction as I’m making here.
represented by “pure qualia” views, eschews any such ambition. They either hold that perceptual character is fundamental and not reducible or explainable by anything else or else are holding out for an explanation of qualia that is not in terms of perceptual relations but instead in terms of something else, e.g., neurology. Since naive realism, as well as the best-known competitor, intentionalism, are paradigmatically relationalist approaches to phenomenal character, it is relationalism that is the focus of this dissertation.

3.1.2 Disjunctivism versus Common Factor Theories

There are two different approaches to theorizing about perceptual experience within relationalism. Consider the following principle:

(CF-Principle) All perceptual experiences have a fundamental factor in common.

For current purposes, a \textit{fundamental factor} is defined as any positive property involving a single fundamental perceptual relation to the same type of ontological entity. Here I’m

\footnote{Papineau (2021) is an example of a non-relationalist. Also, see Block (2010, 2019).}

\footnote{The focus in this dissertation is on theories according to which the fundamental perceptual relation is binary. But there is also the option of a ternary approach. Following Alston (1999) you might contend that the fundamental perceptual relation is a three-place relation between a thing, a property, and a subject (roughly: \(x\) appears \(F\) to \(S\)). Note that for every instance of the ternary relation there is a binary relation to a proposition: Let say you \(R\) the proposition that \(Fx\) if and only if you stand in the ternary relation to \(x\) and \(F\). So, the battle will then be fought over which is more fundamental, the binary relation to propositions or the ternary relation itself. I don’t really see how to make the case that the ternary relation is more fundamental than the binary one. Since the proposition that one \(R\)’s the proposition that \(Fx\) is necessarily equivalent to the proposition that one stands in the fundamental ternary relation to \(F\) and \(x\), the distinction between this approach and intentionalism is very delicate indeed and turns on judgments of relative fundamentality that I do not wish to pursue further here. The debate somewhat recapitulates that between the standard view of singular thought as a relation between a subject and a proposition and Russell’s relational view which treated judgment as a three-place relation between a subject, an object and a property.}

\footnote{If we allow negative properties, we could make a case that \textit{being not knowably distinct from some veridical perception on the basis of reflection alone} counts as a fundamental factor, thereby miscategorizing Martin’s brand of disjunctivism as a common factor theory. The account give here may seem \textit{ad hoc}, but I’m not sure how else to do this without adding much more complexity than seems useful at this point (cf. Pautz 2010: 263-264).}
using fundamental as introduced above (1.4.1) and to be a fundamental perceptual relation is to be one of the most fundamental relations among all the perceptual properties and relations. The sense-datum theory and intentionalism accept the CF-principle since they both hold that all perceptual experience involves a fundamental perceptual relation to the same kind of entity. For the sense-datum theory, all perceptual experiences involve standing in a fundamental relation to sense-data. To be clear, in this example, if \( R \) is the fundamental perceptual relation for the sense-datum theory, the property of \( R \)-ing some sense-datum or other counts as a fundamental factor by the definition above. For intentionalism, all experiences involve standing in fundamental relation to a proposition. So, the property \( R \)-ing some proposition will count as the fundamental factor. Let’s call these common factor relationalist theories, or for short, common factor theories. The CF-principle (or some related version) might be best understood as a default starting position when theorizing about perception, ultimately motivated by the theoretical virtue of simplicity.

We should be clear that the common factor theories say far more than the claim that perceptual experiences have some property in common. That would be a rather trivial claim, assuming that the predicate ‘being a perceptual experience’ expresses a property. Meanwhile the claim that perceptual experiences have a fundamental property in common that is distinctive to perceptual experiences is arguably too strong.\(^77\) Suppose the perceptual relation \( R \) is a fundamental relation. It hardly follows that the property of \( R \)-ing some \( K \) or other (for some chosen \( K \)) counts as a fundamental property. Analogously, if

\(^77\) Note, that if I didn’t require a common property (factor) that is distinctive, it might be too easy to find a fundamental property in common. For example, perhaps there is a fundamental property of being an event that is common to experiences.
the relation _being five feet from_ were a fundamental relation, it hardly follows that _being five feet from a fence_ counts as a fundamental property.

We can distinguish common factory theories from _disjunctivist relationalist accounts_, or for short, _disjunctivist theories_ of perception. Disjunctivists reject the CF-principle in providing a different account of veridical perception and hallucination. This is a weak notion of disjunctivism since a view counts as disjunctivist if it offers a different account of veridical perception and hallucination in terms of either (1) a difference in the fundamental perceiving relation, or (2) a difference in the type of ontological entity.78 To be clear, I will use ‘disjunctivism’ to pick out philosophical theories about the nature of perceptual experience.79 Further, for simplicity, I will assume all disjunctivists are naive realists about veridical perception. Historically, disjunctivism about the nature of perceptual experience has been a defensive strategy to insulate naive realism about veridical perception from the arguments from hallucination.80 I’ll be evaluating this strategy in what follows.

3.1.3 Positive versus Negative Disjunctivism

Within disjunctivism we can make an important further distinction concerning approaches to hallucination. A positive disjunctivist has a positive approach to the character of

---

78 So according to this definition of disjunctivism, even if your theory holds you stand in the same fundamental perceiving relation in both cases of veridical perception and hallucination thereby exhibiting a structural similarity, if the relata at the end of the relation are of different kinds, e.g., tropes versus uninstantiated property complexes, you are still a disjunctivist. Thus, as an example (and to appreciate how I might be deviating from orthodoxy), Johnston (2004) is arguably a disjunctivist on this account.

79 Again, I’m not concerned here with epistemological disjunctivism, a view about differences in rational support or evidence across the categories of experience.

80 Soteriou (2016: 2) essentially makes the same point. Throughout the dissertation, I’ll treat disjunctivism as a form of naive realism.
hallucination, where a positive approach to character is one in which the account of character is given in terms of standing in a (somewhat) fundamental perceptual relation to something or other.\footnote{If cases of hallucination involve a different relation than veridical perception, then the relation in cases of hallucination might be less fundamental than the perceptual awareness relation (involved in veridical perception), but the hallucination relation will still be somewhat fundamental among the perceptual properties and relations.} Note, what makes the positive disjunctivist different from the common factor theory will be one or both of the following: (1) the positive disjunctivist insists that a different perceptual relation is in play between veridical perception and hallucination or (2) the positive disjunctivist allows that the same fundamental relation is in play but posits different ontological kinds of relata at the end of the relation between cases of veridical perception and hallucination.\footnote{There is another type of disjunctivist theory: first, it is a form of naive realism and so relationalist about the character of veridical perception. But this variant is non-relationalist about the character of hallucination. I ignore that option in what follows. First, no one is defending this type of disjunctivism. Second, and more importantly, this type of disjunctivism would be especially vulnerable to the screening off problem.} An example of a positive disjunctivist view is one that holds a subject stands in a relation to a mental object in cases of hallucination (and standing in relation to ordinary objects or tropes in the good case).

The negative disjunctivist denies positive disjunctivism. There are two ways to do this. First, one can be a phenomenal disjunctivist.\footnote{It may be helpful to some readers to point out that while I categorize phenomenal disjunctivism with the term, ‘negative disjunctivism’, Fish categorizes himself as providing a negative epistemic account like Martin, but differing from Martin in providing a “positive account” of hallucination. By positive account, Fish means he provides an account, albeit an eliminativist one about character, rather than no account of hallucination at all. There is no real disagreement between us about the nature of the view, just a difference in taxonomic terminology. Compare Niikawa (2019) who provides a different taxonomy for disjunctivism and puts phenomenal disjunctivism at the top of the decision tree.} The phenomenal disjunctivist says that hallucinations lack character and so there is no character to explain in the case of hallucination. I’ll take up phenomenal disjunctivism below and then again in greater detail in chapter 4. Second, there is the negative disjunctivist who doesn’t deny hallucinatory character. This kind of philosopher wishes to use fundamental perceptual relations to
explain the character of hallucinatory experience but does not wish the explanation to take
the form of explaining the character of a hallucinating subject’s experience at t in terms of
that subject’s standing in a fundamental relation to this that or the other at t. How could a
relation figure in an explanation of the character of a subject’s experience even if the
subject does not stand in the relation to anything? If we use our imagination, we can
concoct all sorts of ways that the relation could still figure in the explanation of
hallucinatory experience without requiring that the subject stand in that relation. One way
is as follows: the subject does not know that they are not standing in a fundamental relation
to such and such. As we shall see this is the evidential disjunctivism advocated by Michael
Martin (2004, 2006, 2013). Another possible way to be a negative disjunctivist: the
subject’s neurology matches a possible subject that stands in a fundamental relation to such
and such. One more example of negative disjunctivism: the subject has beliefs just like it
would if it stood in a fundamental relation to such and such. Having made these three
distinctions in approach, I turn next to the character question.

3.2 A Challenge Concerning the Entities of Perceptual Awareness

Here is the character question for hallucination: what determines the character of a
hallucinatory experience? Many common factor theories are at an advantage in answering
this question since they are able to leverage their fundamental factor. For the sense-datum
theorist, the properties of sense-data determine the character of all perceptual experiences,
including hallucinations. For the intentionalist, which propositions are being perceptually
entertained determines the character of all perceptual experiences, including hallucinations. Can the naive realist take the same common factor approach in answering
the character question? Next, I’ll show there is a significant challenge for a common factor approach to naive realism.

3.2.1 A Challenge for Common Factor Naive Realism

Common factor naive realism is a philosophical theory of perception according to which the core theses of naive realism not only hold for veridical experience but also in cases of hallucination. There is a challenge for common factor naive realism concerning the entities of perceptual awareness. I present the challenge in two steps for a common kind naive realist:

Step One: In order to account for the particularity of experience and the binding problem, naive realism requires that in veridical perception we stand in the relation of perceptual awareness to tropes.

Step Two: Once the common kind naive realist accepts that we stand in relations of perceptual awareness to tropes, then we end up with the counterintuitive result that all hallucination is veridical.

The argument for the first step is presented in chapter 1 (cf. 1.5.4). Next, I’ll present the argument for step two.

For the common factor naive realist, the character of all perceptual experience, including hallucination, is determined by tropes. In cases of hallucination, the subject stands in the perceptual awareness relation to tropes that are instances of properties of internal states, clusters of neurons, photo-recepter cells, etc. Regardless of the details, a
uniform approach to veridical perception and hallucination for the naive realist has a significant cost. Next, for convenience, here are the naive realist bridge principle previously presented (2.4.2):

\[(\forall s)(\forall o)(\forall F) [(a \text{ subject } s \text{ perceives an object } o \text{ to have property } F) \text{ if and only if } (\exists z)(z \text{ is an instance of the property } F \text{ in } o) \land (s \text{ is perceptually aware of } z)]\]

Suppose the common factor naive realist accounts for the character of a hallucination by saying that you perceive trope \(z\). Trope \(z\) will be the instantiation of some property \(F\) by something (be it a sense datum, a neuron, a cluster of neurons) \(o\). Now, \(o\) can only have a trope that is an instance of \(F\)-ness if \(o\) is in fact \(F\). Thus, by the bridge principle you will perceive \(o\) to be \(F\). Similarly, for other tropes that you perceive. The counterintuitive result: all hallucination is veridical.

### 3.2.2 Perceptual Error as Mistakes in Judgement

The common-factor naive realist might learn to live with this conclusion, saying that in hallucination mistakes only arise at the level of judgment and not at the level of perception. All perception as such is veridical but it inclines one to think such things as that there is a pink rat nearby. But I take it this will not appeal to many.

There is an additional problem with the common factor approach to naive realism: veridical hallucination. Take the case of a total veridical hallucination. The judgement will be true but there is total contact failure. Thus, on the view under consideration, there will be no mistake at the level of either perception or judgment. What this view will want to say is that the distinctive feature of veridical hallucination is that the trope perceived
does not explain the truth of the judgment (unlike the good case). Given that both (1) the bridge principle is highly plausible and (2) the counterintuitive results are intolerable, we will have to give up common factor naive realism. Next, I consider disjunctivism.

3.3 The Screening Off Problem

Given the challenge hallucination presents for common factor naive realism, I next turn to the disjunctivist approach. In this section, the focus is on presenting and clarifying the screening off problem for naive realism. The screening off problem is a formidable argument from hallucination targeted against contemporary disjunctivist forms of naive realism. In my view, the screening-off problem has become the new argument from hallucination against naive realism.

In this section, I’ll first present Martin’s original version of the screening off problem (3.3.1). Next, I introduce relevant forms of local supervenience (3.3.2). Finally, I introduce a new version of the screening off problem (3.3.3).

84 There is a related strategy for defending a version of common factor naive realism, called illusionism. Illusionism is the view that all hallucinations are illusions. Here, I’m including those accounts that say all the different sorts of hallucinations can be analyzed as one of the following: veridical perceptions, illusions, or as not perceptual experiences at all. For non-disjunctivist approaches to defending naive realism along these lines, see Ali (2018), Masrour (2020), and Raleigh (2014). This strategy rests on there being a viable naive realist account of illusion. Given my treatment thus far, we might postulate tropes in some region of space in all cases of error. For example, visually hallucinating a snake, gets the same account as misperceiving a rope as a snake. The main idea here is in cases of hallucination (treated as illusions) there is always perceptual contact with a mind-independent entity. Defenders propose contact with an object, instead of a trope. But ultimately, the same challenge remains. The focus shifts back to step 1.
3.3.1 Martin’s screening off problem

In this section, I’m going to present Martin’s screening off problem. The following is my reconstruction of Martin’s original version:

(M1) In cases of veridical perception, one also has the kind of experience that you have in hallucination (call that an H experience).

(M2) Suppose there is a positive account of an H experience including an answer to the character question.

(M3) Given (M1) and (M2), a positive account of an H experience explanatorily screens off the naive realist account of the character of veridical perception, thereby making the account explanatorily redundant.

(M4) If the naive realist account of the character of veridical perception is explanatorily redundant, then naive realism is false.

(M5) Therefore, if the naive realist gives a positive account of H experience, then naive realism is false.

The screening off problem is an argument against positive disjunctivism. Martin takes the conclusion of the screening off problem to highly constrain what the naive realist can say about hallucination, which leads him to his version of negative disjunctivism. There are two general parts to the screening off problem. The first part of the problem is an argument for premise 1. The second part of the problem involves a complex and subtle argument for how having a positive account of the character of hallucination screens off the naive realist account of the character of veridical experience.
Martin (2004: 53-54) provides a detailed argument for the first premise above. It is a unique causal argument from hallucination. Let’s call it the reverse causal argument. Here it is:

(R1) When a subject, s, sees a pine tree at t, call this situation v, there is in s’s body some complete causal condition just prior to t which determined the chance of this event of seeing occurring in v, call this condition N.

(R2) It is nomologically possible that N should occur in s even if no candidate object of perception is present and conditions necessary for the occurrence of a perception are not met, and an hallucinatory experience instead occurs; call one such situation h.

(R3) Where two situations involve the same proximate causal conditions, and do not differ in any non-causal conditions for the occurrence of some kind of effect, then the chances for the occurrence of such an effect are the same in both situations.

(R4) No non-causal condition required for the occurrence of the effects of N is present in h but absent in v.

(R5) Whatever kind of experience occurs in h, there is the same chance of such an experience occurring in v.

(R6) Hence, whatever kind of experience does occur in situations like h, it is possible that such a kind of experience occurs when one is veridically perceiving.

I’m going to be relatively brief in my commentary on the reverse causal argument, focusing mostly on my reasons for providing a different version. First, premise R3 is false
in full generality. Two situations may involve the same proximate causal conditions but one causes the first running of a four minute mile and the second does not. Suppose the second situation comes after the first and the second running is slower than the first running. So, it is not true for all properties $F$, that if two situations have the same proximate causal conditions and the first causes an event that is an $F$, the second causes an event that is an $F$.

There may also be other worries with Martin’s reverse causal argument. First, there may be worries about indeterminism since matching proximate causal conditions can cause different effects in an indeterministic world. Relatedly, there are questions of how chances are being used in the argument. It is likely that Martin is assuming there is no action at a distance, for otherwise non-proximate causes may also matter.

As it turns out, there is a much simpler way to cut through all of this. The key intuition behind the reverse causal argument is that the character of hallucinatory experience supervenes locally on the subject. All the various subtle issues concerning proximate causes are in fact neither here nor there. Below, I will be recasting this argument directly in terms of narrowness which avoids many of these issues. I will introduce narrowness in the next section (3.3.2). Before I do, I want to explore the second part of the screening off problem.

The second part of the screening off problem involves a screening off principle implicit in M3. Here is Martin on the principle at the heart of the problem:

The phenomenon which are in common [e.g., how things appear to the subject, cognitive effects] between the hallucination and the [veridical] perception are accompanied by a common kind of occurrence in both situations. So...those phenomena will have a common explanation in the two situations, namely the occurrence of a kind of experience [re: an H-experience] common to both perception and hallucination and the kind of event which is unique to [veridical]
perceptual situations will be explanatorily redundant. ...the common kind of event between hallucination and [veridical] perception seems better correlated with these common phenomena than the kind of event unique to [veridical] perception and so seems to screen off the purely perceptual kind of event from giving us an explanation. (2004: 61-62, brackets mine)

Here is the principle schematically: Suppose event $e$ has property $H$ and property $V$. In such cases, $e$’s being $H$, screens off $e$’s being $V$ from explaining some effect of $e$, if $e$ had been $H$ and not $V$, the effect would still have occurred (cf. Nudds 2009: 338). It can be difficult to follow Martin’s presentation given the generality in his approach and the reference to kinds. What Martin is ultimately worried about is the status of the naive realist account of the character of experience in veridical perception. The worry is that in explaining how things appear to a subject in cases of veridical perception, the positive account of character of H-experiences (present in both veridical perception and hallucination) will screen off the naive realist account. Since this is the exact issue, I’m going to recast the argument with a razor sharp focus on how a property of an event explains how things appear to a subject. Next, I’ll turn to narrowness.

3.3.2 Narrowness

Intuitions supporting the following causal principles, at least as used in various arguments from hallucination, are arguably not intuitions about causation:

(CP1) It is necessary to give the same account of a veridical perception and a matching hallucination when they have the same proximate neural cause.
(CP2) Where two situations involve the same proximate causal conditions, and do not differ in any non-causal conditions for the occurrence of some kind of effect, then they have the same effect.

Robinson (1994: 151) uses CP1 in his rendition of the causal argument from hallucination. In slogan form, the principle is roughly: *if there is the same proximate cause, then there will be the same effect.* The naive realist rejects this principle, since there is a non-causal condition in cases of veridical perception (i.e., standing in the fundamental perceptual relation). But CP2, a simplified version of Martin’s R3, accommodates the non-causal condition while still relying on the same causal principle. I’ve raised various problems for these general principles in the previous section (3.3.1).

I contend that what is really behind the appeal to causation are intuitions about the local supervenience of character or what I will call *narrowness.* Narrowness about character in slogan form is this: *intrinsic duplication with and across worlds begets phenomenal duplication.* In other words, the properties that specify phenomenal properties are intrinsic. Here is a definition:

A positive phenomenon property \( G \) is narrow if and only if it satisfies the following condition:

Necessarily, for any subject, \( S \), and any complete intrinsic profile \( \varphi \), if \( S \) has \( \varphi \) and undergoes an experience with \( G \), then necessarily, any subject with \( \varphi \) also undergoes an experience with \( G \).\(^{85}\)

\(^{85}\) In formulating this definition I have followed the general approach of Yli-Vakkuri and Hawthorne (2018: chapter 1) but adapted it to phenomenal character and simplified it to suit the purposes at hand.
An intrinsic profile of a subject specifies the complete history of goings-on in the subject—it is a complicated property, e.g., *being a two second long point particle that is green for the first second and blue for the second second and... on and on.* Note, since character is itself just a conjunctive positive phenomenal property (axiom 6), this definition also works for narrow character.

There is a related notion of a *narrow property* that I want to briefly make clear as it will feature in the new version of the screening off problem.

A property $F$ is *narrow* if and only if it satisfies the following condition:

Necessarily, for any subject, $S$, and any complete intrinsic profile $\varphi$, if $s$ has $\varphi$ and has property $F$, then necessarily, any subject with $\varphi$ also has $F$.

This second notion of a narrow property applies to properties of a subject, whereas the first notion of a narrow phenomenal property is a property of an experience. If a property $F$ is a narrow phenomenal property, then the property of having an experience with $F$ is a narrow property. Presumably, if narrow phenomenal properties of experiences supervene on neurological properties, then those neurological properties are also narrow properties.

The main idea behind some of the causal arguments from hallucination, like the reverse causal argument, is that hallucinations have narrow phenomenal properties. Take a case of a completely blind person, Ben, with Charles Bonnet syndrome who experiences a total hallucination. Suppose Ben hallucinates a person walking toward him. There is a strong intuition that in such cases of total hallucination, the character of experience is an entirely internal matter— that the internal state and internal history of the subject determine
the character. I use the example of Charles Bonnet because such cases occur in the actual world. But there are other possible cases: my perfect hallucination of my backyard while testing ReViz in a pitch-black room, non-veridical perceptual experiences directly following sudden envatment, and Cartesian demons causing total hallucinations. In all of these actual and possible cases of total hallucination, the character of these experiences appear entirely dependent on the internal state and history of the subject, and so not determined by the subject standing in fundamental perceptual relations to mind-independent concreta (or their tropes). Next, I’ll recast the screening off problem using narrowness instead of relying on a causal principle.

3.3.3 A New Version of the Screening Off Problem

In this section, I’ll present a new version of the screening off problem. But before I do, I need to first introduce a special kind of phenomenal property, a naive realist phenomenal property (or NR-phenomenal property for short). Here is a definition:

A phenomenal property being $F$ counts as an NR-phenomenal property if and only if an experience has $F$ in virtue of satisfying the core theses of naive realism.

For the naive realist, a veridical perceptual experience has its phenomenal character at least partly in virtue of the subject standing in fundamental relations of perceptual awareness to mind-independent entities that are constituents of the experience. Since character is just the conjunction of the phenomenal properties of the experience, both the character and some of its conjuncts are NR-phenomenal properties.
The new version of the screening off problem has a different structure than the original. Here is the new screening off problem in the form of a deductive argument:

(N1) A subject, $S_h$, undergoes a total hallucinatory experience with phenomenal property $P$ that explains how things appear to $S_h$. (For simplicity, $P$ is the only phenomenal property of the experience and so $P$ is also the character of the hallucinatory experience.)

(N2) The character of a total hallucinatory experience is narrow.

(N3) A subject, $S_d$, is an intrinsic duplicate of $S_h$, and $S_d$ is undergoing a veridical perceptual experience such that things appear the same way to $S_h$ and $S_d$.

(N4) Since $S_d$ is an intrinsic duplicate of $S_h$ and $P$ is narrow, $S_d$ also undergoes an experience with $P$.

(N5) If $S_h$ and $S_d$ both undergo an experience with $P$ and things appear the same way to $S_h$ and $S_d$ and $P$ explains how things appear to $S_h$, then $P$ explains how things appear to $S_d$.

(N6) From N1, N3, N4, and N5, $P$ explains how things appear to $S_d$, which is a case of veridical perception.

(N7) Since NR-phenomenal properties are not instantiated in cases of total hallucination, they play no explanatory role in cases of total hallucination.

(N8) In cases of total hallucination, it is (if anything) only narrow properties that explains the character of experience, and so it is only narrow properties that explain $P$ in the case of $S_h$. 
If narrow properties explain $P$ in cases of total hallucination and the same narrow properties are present in a case of veridical perception, then the narrow properties explain $P$ in the case of veridical perception.

From N4, N8 and N9, narrow properties explain $P$ in the case of $S_d$. And so narrow properties explain how things appear to $S_d$, which is a case of veridical perception.

If there is no explanatory overdetermination in veridical perception, then NR-phenomenal properties do not explain how things appear to $S_d$, which is a case of veridical perception.

There is no explanatory overdetermination of how things appear in cases of veridical perception.

From (N11) and (N12), NR-phenomenal properties do not explain how things appear $S_d$, which is a case of veridical perception.

Conclusion: Naive realism is false.\footnote{The new version of the screening off argument in the main text is long-winded. Perhaps there is a shorter way to spell it out without loss of clarity. In a previous attempt to construct a new version of the screening off argument using narrowness, I had a quick-and-dirty version that was the same premises N1-N6 as above, but then differed by finishing this way:

(N7*) NR-phenomenal properties do not explain why any experience has $P$, since a total hallucination can have $P$.

(N8*) From N6 and N7, NR-phenomenal properties do not explain how things appear to $S_d$.

The argument ended with N8* which shows naive realism is false. John Hawthorne pointed out to me a problem with the argument that I had missed. Just to be clear, the argument through premise N6 already has the naive realist in a tight spot. But an argument is needed from N6 to claiming naive realism is false. The problem with the quick-and-dirty version is this: N8* does not follow from N6 and N7*. In order to illustrate the problem, I’ll present another instance of the argument form where the premises are true and the conclusion false:

(B1) Having the ability to create lift explains how birds can fly.
(B2) Having feathers does not explain why any animal is able to create lift, since bats are able to create lift.
(B3) Therefore, having feathers does not explain how birds can fly.}

(N10) From N4, N8 and N9, narrow properties explain $P$ in the case of $S_d$. And so narrow properties explain how things appear to $S_d$, which is a case of veridical perception.

(N11) If there is no explanatory overdetermination in veridical perception, then NR-phenomenal properties do not explain how things appear to $S_d$, which is a case of veridical perception.

(N12) There is no explanatory overdetermination of how things appear in cases of veridical perception.

(N13) From (N11) and (N12), NR-phenomenal properties do not explain how things appear $S_d$, which is a case of veridical perception.

Conclusion: Naive realism is false.\footnote{The new version of the screening off argument in the main text is long-winded. Perhaps there is a shorter way to spell it out without loss of clarity. In a previous attempt to construct a new version of the screening off argument using narrowness, I had a quick-and-dirty version that was the same premises N1-N6 as above, but then differed by finishing this way:

(N7*) NR-phenomenal properties do not explain why any experience has $P$, since a total hallucination can have $P$.

(N8*) From N6 and N7, NR-phenomenal properties do not explain how things appear to $S_d$.

The argument ended with N8* which shows naive realism is false. John Hawthorne pointed out to me a problem with the argument that I had missed. Just to be clear, the argument through premise N6 already has the naive realist in a tight spot. But an argument is needed from N6 to claiming naive realism is false. The problem with the quick-and-dirty version is this: N8* does not follow from N6 and N7*. In order to illustrate the problem, I’ll present another instance of the argument form where the premises are true and the conclusion false:

(B1) Having the ability to create lift explains how birds can fly.
(B2) Having feathers does not explain why any animal is able to create lift, since bats are able to create lift.
(B3) Therefore, having feathers does not explain how birds can fly.}
Note, as advertised, there is no more need for a reverse causal argument or appeal to proximate causes in this version of the screening off argument. There is also no appeal to ‘kinds’ in the argument. The conclusions of the original and new version of the problem are also different. The conclusion of the original version of the argument makes positive disjunctivism an unstable position and so is an argument against positive disjunctivism. The conclusion of the new version renders NR-phenomenal properties explanatorily impotent. So, the new version of the problem targets all naive realists (both positive and negative disjunctivists). Any naive realist will have to deny one or more of the premises to defend their view. Moving forward, when I use the expression, ‘screening off problem’, unmodified, I’m referring to the new version which is simply the argument made up of premises N1 to N13. Next, I’ll examine replies.

3.4 Replies to the Screening Off Problem

In this section, I’ll evaluate various replies to the screening off problem. First, I’ll consider replies on behalf of the two most developed forms of negative disjunctivism: evidential disjunctivism (3.4.1) and phenomenal disjunctivism (3.4.2). Next, I’ll look at various replies on behalf of positive disjunctivism that involve denying narrowness (3.4.3). Lastly, I’ll present a novel reply to the screening off problem (3.4.4).

---

Even if validity is restored by adding a conditional premise cloaked in ‘explains’ talk, i.e., if B1 and B2, then B3, the conditional premise is false. It’s easy to spot the problem in the bird argument. Thanks to Lisa Kovakovich for the example. For whatever reason, I found the flaw harder to spot in the original: If P is not an NR-phenomenal property and P explains how things appear to Sx while undergoing a veridical perceptual experience, then NR-phenomenal properties do not explain how things look to subjects in veridical perception. Nevertheless, the naive realist will simply argue that the conditional is false.
3.4.1 Evidential Disjunctivism

Arguably, no contemporary philosopher has been more influential in reviving naive realism than Michael Martin (2004, 2006, 2013). Martin defends a form of negative disjunctivism, he calls *evidential disjunctivism*. Next, I present two direct quotes from Martin that make up the key claims of evidential disjunctivism:

(ED1) The notion of a visual experience of a white picket fence is that of a situation being indiscriminable through reflection from a veridical visual perception of a white picket fence as what it is. (2004: 363)

(ED2) For certain visual experiences as of a white picket fence, namely, causally matching hallucinations, there is no more to the phenomenal character of such experiences than that of being indiscriminable from corresponding visual perceptions of a white picket fence as what it is. (2004: 369)

ED1 and ED2 both employ the notion of being indiscriminable from a veridical perception. For Martin, the notion is a negative property, roughly: *not being knowable through reflection alone from a certain veridical perception*. In other words, it is not possible for the subject to know through reflection alone that they are not having a veridical perception of such and such. This negative property includes reference to the naive realist account of veridical perception as specified by the core theses. Further, in ED1 this negative property is used to fix the extension of what counts as a certain visual experience. In ED2, the negative property is used to provide a negative account of hallucination.
Given ED2, Martin will deny premise N7 in the screening off problem. According to his view, NR-phenomenal properties still play a role in the explanation of how things look in hallucination since the negative account of the character of hallucination is indiscriminable from a veridical experience with NR-phenomenal properties.

Next, I’ll present an argument supporting premise N7 which is also incompatible with evidential disjunctivism. It makes use of the following privation principle:

\[(\forall x)(\forall F)((F \text{ is a positive property } \land Fx) \rightarrow \neg(\exists G)(\neg Gx \land G \text{ explains the fact that } Fx)).\]

The privation principle states that an entity having a positive property is never to be explained by a property that entity lacks. This principle is highly intuitive. Here is the argument from privation against evidential disjunctivism:

1. A subject undergoes a total hallucinatory experience, \(h\), with positive phenomenal property \(P\).
2. By the privation principle, that \(h\) has \(P\) is not to be explained by any properties \(h\) lacks.
3. Since \(h\) is a total hallucination, \(h\) lacks NR-phenomenal properties.
4. Therefore, that \(h\) has \(P\) is not to be explained by NR-phenomenal properties.

Obviously, the conclusion is incompatible with Martin’s evidential disjunctivism. Martin has two options remaining: deny phenomenal properties are positive properties or deny the privation principle. I suspect Martin will contend that some phenomenal properties are negative. I gave several reasons for holding all phenomenal properties are positive in
chapter 1 (1.3.3). Many will find neither defending negative phenomenal properties nor denying the privation principle tolerable.

I’ve just presented a novel argument against Martin’s evidential disjunctivism in the context of his preferred solution to the screening off problem. In addition, several serious objections have been raised against evidential disjunctivism (Byrne and Logue 2008; Farkas 2006; Hawthorne and Kovakovich 2006; Pautz 2010, 2011; Seigal 2004, 2008). Since I have already published a paper (with Hawthorne) cited in the main text that is focused on the other problems with evidential disjunctivism and there has been no counter-response, I’m not rehashing those arguments here. Given the argument from privation and the many other substantial arguments against evidential disjunctivism, it is likely not the most viable way forward for the naive realist. Next, I move on to considering the other form of negative disjunctivism.

3.4.2 Phenomenal Disjunctivism

Phenomenal disjunctivism is the view that total hallucinations have no phenomenal properties. William Fish (2008; 2009; 2013b), Heather Logue (2013), and Scott Sturgeon (2006; 2008) have each argued for a form of phenomenal disjunctivism.\(^8\) This is a radical approach to the screening off problem that denies premise N1. However, the screening off argument can be amended to make trouble with at least one version of phenomenal disjunctivism (a revenge screening off problem). Before I turn to the revenge screening off problem, I first want to address two salient issues for phenomenal disjunctivism.

---

\(^8\) Fish is the only philosopher of the three that develops and defends phenomenal disjunctivism in detail. Thus, my focus on his account of the view throughout the dissertation.
The first salient issue with phenomenal disjunctivism is the proposed error theory of hallucination. In pre-theoretically reflecting on the phenomenology of hallucinatory visual experiences, the following seems true: total hallucinatory visual experiences (and hallucinatory elements of a partial visual hallucination) have visual phenomenal properties. The phenomenal disjunctivist claims this pre-theoretical finding is in error. It does appear there is methodological double standard in play by the phenomenal disjunctivist leaning so heavily on pre-theoretical reflection to motivate their account of veridical perception (see chapter 2) while at the same time dismissing the deliverances of pre-theoretical reflection in their account of hallucination. Ideally, any such error theory will also provide an explanation of why we are mistaken. As I will unpack below, Fish (2009) does provide such an explanation. Nevertheless, for many (especially those who have experienced hallucinations) the fact that the theory is claiming that total hallucinators are phenomenal zombies (at least as far as the perceptual aspects of experience goes) already makes the theory completely untenable.

The second salient issue with phenomenal disjunctivism is that the view conflicts with axiom 2 of the axioms of phenomenology (1.3.1). Axiom 2 defined an experience as an event instantiating at least one positive phenomenal property. On this view, all first-order phenomenal properties of perceptual experience are NR-phenomenal properties and total hallucinations lack NR-phenomenal properties. Thus, a total hallucination is, by definition, not an experience. The phenomenal disjunctivist will likely hold that total hallucinations are a different kind of experience-like mental event. They will then owe us an account of why they are grouped with the traditional categories of perceptual experience and also an account that differentiates them from unconscious mental states. They will
also need to provide another account of phenomenology that allows for conscious experience consistent with their position on hallucination, since it is undeniable that hallucinations have a felt reality.

Alternatively, a phenomenal disjunctivist might be tempted to back off the strong claim that total hallucinations have no phenomenal properties whatsoever and instead claim that total hallucinations only have those phenomenal properties of contemporaneous cognitive experiences involving perceptual judgments. The problem with this approach is that if hallucinations have any kind of phenomenal properties that explain how things look to the subject, then this type of disjunctivism does not yet have a solution to the screening off problem, since they no longer can deny premise N1. For this reason, the phenomenal disjunctivist will deny that hallucinations have any phenomenal properties but instead provide an alternative account of their felt reality.

In the next chapter, I present Fish’s account of hallucination in detail. In this section, I will only provide a rough sketch of the view in order to evaluate the viability of this response to the screening off problem. According to Fish’s account, a total visual hallucination of an F is (roughly) a mental event, lacking phenomenal properties, that has the same cognitive effects as a possible veridical visual perception of an F: Cognitive effects are perceptual judgments and beliefs about what the subject takes themself to be seeing. Fish appeals to the cognitive effects to explain the felt reality of hallucination. He explains:

A mental event both becomes a hallucination and acquires a felt reality as it has its cognitive effects. If the effects do not occur, then the event will not attain a felt reality, nor will qualify...as a hallucination. (2009: 100)

88 The hallucinating subject and their veridical perceiving counterpart are required to share background beliefs. For further details see chapter 4.
...what it is like to hallucinate is determined by the kind of veridical perception one mistakenly takes oneself to be enjoying. ...whatever hallucinations contribute to what it is like for a conscious subject, both the very fact that they contribute at all, as well as the particular contribution they make, is entirely parasitic on the contribution made by veridical experience. (2013b: 64)

Fish’s account of the felt reality of hallucinations can be taken as answering the two salient issues for phenomenal disjunctivism raised above. First, the account of felt reality can be taken as an explanation for why it seems total hallucinations have phenomenal properties. Still, there appears to be a gap between (1) the felt reality of visually hallucinating a dragon and (2) having the belief that one sees a dragon. Second, Fish will have to provide an alternative account of conscious experience and phenomenology (than I have provided here) such that hallucinations can contribute to “what it is like for a conscious subject” without having phenomenal properties.

There is a related challenge in the vicinity concerning partial hallucinations for Fish. According to Fish, in a partial hallucination, the hallucinatory element embedded in a veridical perceived background is given a similar account in terms of having the same cognitive effects as a matching veridical experience. Suppose you are in your living room suffering a partial hallucination of a giraffe sitting on the couch. You are hallucinating the giraffe while simultaneously veridically perceiving the couch and other ordinary items in view. Here is the challenge for Fish: What is it like to visually experience a hallucinated giraffe sitting on my veridically perceived couch in my living room? How does Fish stitch together the character of veridical experiences and the lack of character of hallucination to explain the character of partial hallucination?89

89 Ted Sider raised this concern regarding Fish’s view to me.
While Fish doesn't explicitly address this challenge, I think he has two responses available. One response is to say strictly speaking that the character of your visual experience is the same as if you were not having the partial hallucination. So, your visual field includes the entire empty couch, without any giraffe features. Your beliefs about your visual experience to the contrary. The second (and perhaps more plausible) response available to Fish is to say that the visual phenomenal properties of the portion of your visual field where the partial hallucination is occurring is phenomenally like a temporary scotoma—which, for current purposes, can be characterized as the absence of visual phenomenology in a specific region while also lacking any awareness of said absence. I’m not sure these responses are ultimately satisfying. But any residual discomfort is likely due to the error theory. Having established that eliminativism about the character of hallucination comes with a price, how does it solve the screening off problem?

If total hallucinations lack phenomenal character, then there is no narrow property there to screen off the explanatory work of NR-phenomenal properties. But for the solution to stick, Fish will have to be thoroughly eliminativist about how things appear. Suppose someone visually hallucinates a giraffe. If there is a total visual hallucination, phenomenal disjunctivism denies that the visual experience has phenomenal properties. But what about predicates like ‘looks to S as if there is a giraffe’ and the property expressed by that predicate? A mad dog error theory would have an error theory here too. They would say that it doesn’t look to the hallucinator as if there is a giraffe and that the hallucinator merely thinks that it looks to the hallucinator as if there is a giraffe. (Call this view mad dog disjunctivism.) Now suppose a phenomenal disjunctivist doesn’t want to be a mad dog disjunctivist (since the mad dog view is especially crazy). Rather, this other kind of
phenomenal disjunctivist would say what makes it true that it looks to the hallucinator as if there is a giraffe is the hallucinator is experiencing the same cognitive effects that a suitable counterpart experiences when seeing a giraffe. But now there is a new version of the screening off problem for the property of looking as if there is a giraffe. The property of having the same cognitive effects that a suitable counterpart experiences when seeing a giraffe explains why it looks as if there is a giraffe when you see a giraffe. The NR-phenomenal properties do not explain it looking as if there is a giraffe even in cases when you see a giraffe.

If Fish doesn’t embrace mad dog disjunctivism, he’ll end up facing another version of the screening off problem. In which case, he will likely opt for the same route out of the screening off problem as Martin and deny N7. Now, as I have shown, that route is not a promising one. It remains unclear from his published work if Fish is a mad dog disjunctivist or not. I’ll return to Fish’s evidential disjunctivism in the next chapter. Next, I’ll move on to considering positive disjunctivist replies to the screening off problem.

3.4.3 Positive Disjunctivism

According to positive disjunctivism, the character of hallucination is determined by the entities to which we stand in fundamental perceptual relations. But the account of the character of hallucination is different from the account of the character of veridical perception as to avoid the challenges presented to common factor naive realism (3.2). For purposes of illustration, I’m going to sketch one potential way to be a positive disjunctivist with a few placeholders for generality. In hallucination, either the fundamental perceptual relations or the entities serving as relata or both must be different from veridical perception.
One promising route forward is to hold that the entities of awareness in hallucination are *mind-dependent* entities (while maintaining that in veridical perception, the entities of awareness are mind-independent entities). Since minds have the capacity to create entities of “awareness” for purposes of visualizing, dreaming, imagining, etc., those entities are thus mind-dependent. Such entities might be mental images, sense-data, or mind-dependent tropes. On this view, the character of hallucination is determined by the particular mind-dependent entities of which the subject is standing in fundamental perceptual relations. I’ll call this type of disjunctivism, *MD-positive disjunctivism*.

I have left open whether the fundamental perceptual relation in the case of hallucination is the same or different. One might think that two relations are distinct if the ontological category of their relata are different. Also, given the directness thesis there is some pressure to posit a distinct perceptual relation obtaining in cases of hallucination. Otherwise, the possibility that the mind-dependent entities mediate our awareness to mind-independent entities is a looming threat. Let’s suppose it is a different fundamental perceptual relation, call it *quasi-perceptual awareness*. So put this all together, we get the following two theses of total hallucination for MD-positive disjunctivism:

**TH-Ontological Thesis:** In total hallucination, a subject stands in the relation of *quasi-perceptual awareness* to mind-dependent entities, and this relation is (somewhat) fundamental among the perceptual properties and relations.

**TH-Phenomenological Thesis:** In total hallucination, the character of experience is determined by the particular mind-dependent entities to which the subject stands in *quasi-perceptual awareness*. 
‘Quasi-perceptual awareness’ is a label for the (somewhat) fundamental perceptual relation that holds in cases of total hallucination. Hallucination might be a degenerate form of perception or it might be a sui generis mental activity (in which case, it might be better labeled ‘hallucinatory awareness’) that falls within general perception or it. This account of total hallucination is also compatible with the position that hallucination is a degenerate form of sensory imagination (Allen 2015). With these positive disjunctivist core theses of total hallucination in place, I now turn to potential replies to the screening off problem.

The positive disjunctivist might deny N2, that the phenomenal character of a total hallucination is narrow. This is going to require positive disjunctivism to provide more of an account. To see why, first recall that total hallucinations are cases that involve no perceptual contact with entities in the environment of the subject, i.e., the outer world. Since there is no perceptual contact with the outer world, the experience is only dependent on the intrinsic profile of the subject. According to the positive disjunctivist view under consideration, the character of a total hallucination is determined by which mind-dependent entities you stand in relations of quasi-perceptual awareness. Arguably, quasi-perceptual relations to mind-dependent entities are entailed by the intrinsic profile of the subject whose mind the entities are dependent on. But if that were the case the character would be narrow. So, to deny N2, one will either have to say that the mind-dependent entities do not supervene on intrinsic profile or else say that the obtaining of the quasi-perceptual relations do not supervene on intrinsic profile. And in particular, one who wants to block the screening off problem by this kind of move will want to say that in veridical perception one does not stand in the relevant quasi-perceptual awareness relations to the relevant mind-dependent entities.
One proposal is that the quasi-perceptual awareness relation to mind-dependent entities only obtains in the absence of a suitable mind-independent entity in the environment of the subject. Another option is to hold that the mind-dependent entities only exist when there is no suitable mind-independent entity in the environment of the subject. One problem with this approach is that it is unable to account for total veridical hallucination. This approach requires that all total hallucinations (including veridical hallucinations) are object-failures, in addition to being contact failures (1.2.1). But a total veridical hallucination is an experience with total contact failure but is object-successful. Intuitively, cases of veridical hallucination turn on an implicit counterfactual in the background: that the subject would be suffering the same hallucination even if the suitable objects weren’t present in the environment. We might attempt to accommodate these cases by amending the condition with a causal constraint: roughly, that the experience lacks the property of being (partly) caused by a suitable object in the environment. But now if it is a property an experience lacks that explains a phenomenal property, then we run up against the privation argument again. So, perhaps this approach will also deny that the character of hallucination is a positive property.

---

90 Allen (2015) advances a view in this spirit, however he is providing a solution to the original screening off problem where it is a kind of experience playing center stage (an H experience) rather than the character of hallucination. The bigger problem for this view and Allen’s is not just veridical hallucination, but is a question concerning what is going on in the mind of intrinsic duplicates across good and bad cases: is the mind-dependent entity present or not in the good case? Even if Allen says the hallucinatory event doesn’t supervene on the brain in veridical perception because there is a suitable external object present, there is still the question regarding the presence or absence of whatever explains the character of hallucination.

91 Moran (2019) provides a solution to Martin’s original screening off problem by proposing causal conditions on kinds of experience and then using those conditions to deny that experiences of the hallucinatory kind supervene locally on their subject’s brain. Moran’s proposal does not answer the new version of the screening off argument presented here. He maintains that phenomenal properties are narrow in stating, “…subjects in the same brain state might differ in terms of what kind of experience they have. …Indeed, sameness of brain state guarantees sameness of phenomenology: all that differs depending on the causal context is the type of experience that the subject has” (2015: 377, fn 23). Thus, it is not clear he would adopt a causal condition on character as I’m entertaining here.
Let’s be clear about just how zany this approach is turning out. The view is thoroughgoing relationalist: in cases of total hallucination, similar to cases of veridical perception, the character of the experience is explained by its subject standing in a somewhat fundamental perceptual relation to some entities. But they deny that the existence of those entities and/or the obtaining of that relation supervenes on narrow aspects of the subject, even when the subject is floating along in a void. This is odd.

I want to make a related point here just to be clear. When I ask for an explanation of the character of hallucination, I am asking about what relational facts make for the characters that obtain in cases of hallucination. There is a different question. What makes a case count as a case of hallucination? Now of course the absence of suitable relations to mind-dependent entities in the environment can explain why the predicate ‘is a hallucination’ applies to a case. But it is quite another matter to claim that the phenomenal properties in those cases are to be explained by entities and relations that exist and obtain only thanks to the fact that the case is a hallucination.

Denying N2 is counter-intuitive. Many will find its denial a price they are unwilling to pay. The MD-positive disjunctivist account is of no help. Considering the approach as I have shows just how radical it is for an MD-disjunctivist to deny N2. Next, I’ll consider a closely related reply, denying N8.

Suppose the MD-positive disjunctivist accepts N2, conceding that whatever mind-dependent entities and quasi-perceptual relations that are present in the hallucinatory case are also present in cases of veridical perception. Then, there is not much prospect for denying N8, since one will then think that narrow properties explain the phenomenal properties that make up the character of a total hallucination. So, the MD-positive
disjunctivist that accepts N2 will find it difficult to deny N8. Next, I’ll present another form of positive disjunctivism for which denying N8 may be more promising.

There is another type of positive disjunctivist better positioned to deny N8. This positive disjunctivist provides the same account of the character of hallucination as MD-positive disjunctivism except for one important difference: the entities to which we stand in quasi-perceptual awareness relations are mind-independent (call this view, MI-positive disjunctivism). One way to be an MI-positive disjunctivist is to hold that in cases of total hallucination, we stand in quasi-perceptual awareness relations to a complex of universals, which are mind-independent abstract entities (cf. Johnston 2004). Now, on this view we may also stand in various relations to complex universals in cases of veridical perception, but we do not stand in the relation of quasi-perceptual awareness and so such quasi-perceptual relations are not present in veridical perception to explain the character of experience. So, this MI-positive disjunctivist can deny that only narrow properties explain the character of total hallucination, at least insofar as they deny that a subject’s standing in relation of quasi-perceptual awareness to an abstract universal across the great line of being is not a narrow property. According to this approach, the quasi-perceptual awareness relation is not instantiated in cases of veridical perception where the fundamental relation of perceptual awareness (to tropes) explains the character of experience. Next, I’ll attempt to motivate the MI-disjunctivist approach.

---

92 Another way to be a MI-positive disjunctivist, is to adopt intentionalism for total hallucination, according to which a subject stands in a (somewhat) fundamental perceptual relation to a proposition. Logue 2013 considers such a view. This view is difficult to maintain since there will be pressure to abandon disjunctivism and adopt common factor intentionalism. And yet another way is to hold that in cases of total hallucination we stand in perceptual relations to meinongian objects (Knight 2014, Parsons 1980).
Here is a toy model of “awareness” for MI-positive disjunctivism. There is a ranking system for what a subject is aware of in a perceptual experience. Here is a simple ranking:

1. Physical objects (or else mind-independent tropes of physical objects)
2. Mind-independent complexes of universals

So, when there is a fence present and there is a suitable causal relation to you, then you are aware of the fence (by standing in the relation of perceptual awareness to the tropes of the fence) and not aware of a complex of universals. Of course, the complex of universals will be around in either case. But we are not aware of the complex of universals in the veridical perception case. Next, I’ll present an analogy with kicking.

Here is a toy model of “kicking”. You kick the highest rank candidate:

1. Baseball
2. An object shaped like the front of a baseball.

So, when there is a baseball present, you kick the baseball but not its front. But when there is only a piece of baseball (shaped like the front of a baseball) and no baseball present, you kick an object shaped like the front of a baseball. On the MI-positivist disjunctivist proposal under consideration, the entities of awareness across various cases of perception are ranked like in the toy models above.

Returning to the screening off problem, this view can consistently maintain that the phenomenal property $P$ is instantiated in both cases, but in neither case does a narrow property explain the character of your experience. Of course, the MI-disjunctivist might fall back and simply deny N2 like the MD-disjunctivist. But as we have seen they can go
along with N2 and deny N8 instead. However, there remains some tension in accepting N2 and denying N8, since you might wonder why the character of a total hallucination is always narrow if what explains the character is not itself a narrow property. I’m unsure if this move (accepting N2 and denying N8) is deeply problematic or is simply a strange combination of views.93

There are problems with MI-disjunctivism. The analogy with the toy models breaks down in cases of veridical hallucination since all we have left to appeal to is the absence or presence of the various fundamental perceptual relations, since in veridical hallucination the fence (or baseball) is present. But the main challenge for the view is that it remains mysterious why the quasi-perceiving relation is uninstantiated in the case of veridical perception. Another challenge is that the view appeals to a complex of abstract universals in its account of character and so it is unclear how it is to account for the particularity of perception. Some may find it unpalatable to put complexes of universals to work in a theory of perception. Finally, it is worth mentioning that a relation determined by a ranking system along the lines of the toy model hardly seems fundamental relative to all properties and relations within the domain of perception. For these reasons, I will not be pursuing MI-positive disjunctivism any further.

There are positive disjunctivists that allow phenomenal property P (the character of the total hallucination in the argument) to be instantiated in cases of veridical perception, where things appear the same way, and other positive disjunctivist that deny any such

93 That said, I have no general argument that narrow properties cannot always be explained by non-narrow properties. Suppose you buy into a standard grounding thought that disjunctive properties are grounded by disjuncts. Consider the property of either being near a llama or not being near a llama. If one buys into that grounding principle, then the disjunctive property is grounded by being near a llama (a non-narrow property) when you are and grounded by not being near a llama (also a non-narrow property) when you are not. But the disjunctive property is narrow since it is shared by all intrinsic duplicates. That said, the current MI-disjunctivist package continues to strike me as very strange.
possibility (by denying N2). For the MD-positive disjunctivist that allows P to be instantiated in cases of veridical perception, there are only two defensible moves left to consider: denying premise N12, thereby allowing for explanatory overdetermination, and denying premise N5. Next, I will briefly consider denying premise N12.

For convenience, here is premise N12: *there is no explanatory overdetermination of how things appear in cases of veridical perception.* The positive disjunctivist who denies N12 allows for P to be instantiated in both the case of total hallucination and the case of veridical perception. In addition to the narrow phenomenal property P, the veridical perception also has NR-phenomenal properties. Call the conjunction of just the NR-phenomenal properties instantiated by the veridical perception which S_d is undergoing, P^*. Call the character of the veridical perception which S_d is undergoing, Q. Given the veridical experience S_d is undergoing does not have any non-narrow phenomenal properties other than NR-phenomenal properties, Q is the conjunction of P and P^*. Call this kind of positive disjunctivism, *Multiplex Positive Disjunctivism.* The character of veridical perception is “multiplex” in the sense that it is always a conjunction of both narrow phenomenal properties and NR-phenomenal properties. The multiplex positive disjunctivist that denies premise N12 maintains that while P explains the way things appear to S_d, Q also explains the way things appear to S_d (cf. Sethi 2020). Further, this move requires that this kind of overdetermination is unproblematic. Suppose we asked whether it is phenomenal property P or phenomenal property Q that explains the way things appear to S_d? This disjunctivist says P explains the way things appear to S_d (premise N6) and that Q explains the way things appear to S_d. This reply, while not logically inconsistent, is highly counterintuitive. Given P is a conjunct of Q, the multiplex positive disjunctivist
might be better off saying that $Q$ (rather than $P$ or both) explains the way things appear to $S_d$. I turn to this reply next.

3.4.4 A Novel Reply

In this section, I will consider one last reply to the screening-off problem on behalf of the multiplex positive disjunctivist. This reply denies premise N5.

(N5) If things appear the same way to $S_h$ and $S_d$ and $S_h$ and $S_d$ both undergo an experience with $P$, and $P$ explains how things appear to $S_h$, then $P$ explains how things appear to $S_d$.

This reply is an option for the multiplex positive disjunctivist. Given (1) things look the same way to $S_h$ and $S_d$, and (2) $S_h$ and $S_d$ both undergo an experience with $P$, and (3) $P$ explains how things appear to $S_h$, this multiplex positive disjunctivist denies that $P$ explains how things appear to $S_d$. Instead, $Q$ explains how things appear to $S_d$. The multiplex positive disjunctivist holds that the character of a perceptual experience explains how things appear to the subject undergoing the experience.\(^{94}\) To further bolster this reply on behalf of multiplex positive disjunctivism, we might put forward the following general principle of explanation:

(Principle E) A thing (or property) that explains is the most natural thing on which there is the appropriate counterfactual dependence.

\(^{94}\) Note it is consistent with this view to hold that distinct characters are indistinguishable.
Let me attempt to both motivate and illustrate the principle with a few toy examples. First, consider these two scenarios:

Scenario 1: A baseball hits a window

Scenario 2: Something shaped like the front half of a baseball hits a window.

Now, suppose we wanted to explain the window breaking event. The object shaped like the front half of a baseball is present in both cases. If there is a front half and a breaking in both cases you might think the front half is explanatorily adequate and hence the baseball is irrelevant. This would be to appeal to the same principle behind premise N5. But that is counterintuitive. When the baseball is present it is the most natural thing to explain the breaking--arguably, window breaking is more dependent on the baseball than the front of the baseball alone. Let’s look at another example. Take these two scenarios:

Scenario 1: Something that is a car collides with a cyclist.

Scenario 2: Something that is a front half of a car collides with a cyclist.

Next, consider what explains a cycle injury event in scenario 1. By the same principle at work in N5, since the front half of a car explains the cycle injury event in scenario 2 and a front half of a car is present in scenario 1, the front half of a car explains the cycle injury event in scenario 1.

If the worry is that there is still explanatory overdetermination in the case of veridical perception since phenomenal property $P$ is still present, this disjunctivist answer is that there is no more overdetermination than in case of baseballs breaking windows and cars injuring cyclist--which according to this view is no overdetermination at all. This response is comfortably combined with both the TH-ontological thesis and TH-
phenomenological thesis. Further, a proponent of the view can also comfortably hold that the character of perceptual experience, while modally determined by fundamental perceptual relations, locally supervenes on intrinsic neurological properties of the subject.

The main problem for the multiplex positive disjunctivist is this: it is counterintuitive that the character of every veridical perception contains as a conjunct a narrow phenomenal property that could explain the way things appear in a subjectively matching case. Call this strange result, the subjectively matching conjunct problem. One case is a subjectively matching case of another if and only if things appear the same to the subject in both cases. To see how this result follows, first note that for every veridical perception there is a possible subjectively matching total hallucination with an intrinsic duplicate. Since the character of a total hallucination is narrow, the phenomenal property that is that character is instantiated by a subjectively matching veridical perception with an intrinsic duplicate. Thus, the character of any veridical perception has a narrow phenomenal property as a conjunct that subjectively matches it but does not explain how things appear other than in cases of total hallucination. One would like more of an account of this strange result before taking up such a view.

3.5 Concluding Remarks

The challenge for common factor naive realism (3.2) makes such a view a non-starter. So, in order to defend a general philosophical theory of perception, the naive realist must adopt disjunctivism, as weakly defined (3.1). The screening off problem is a formidable challenge for disjunctivism (3.3). The good news for the naive realist is that there are many
replies to the screening off problem on offer (3.4). The bad news for the naive realist is that none of the replies is without a price.

If forced to decide between the positive disjunctivist replies, denying premise N5 on behalf of multiplex positive disjunctivism seems to have the least counterintuitive consequences (3.4.4). Based on the considerations and arguments in this chapter, positive disjunctivists that deny N2, N8, or N12 are better off denying N5 instead. This direction shows some promise but requires an explanation for the subjectively matching conjunct problem. While perhaps the most radical reply, Fish’s phenomenal disjunctivism, is the most developed version of negative disjunctivism. I will turn to phenomenal disjunctivism in the next chapter.
Chapter 4
Phenomenal Disjunctivism

The screening off problem is a formidable argument against naive realism. A radical way out of the screening off problem is to adopt phenomenal disjunctivism which denies that total hallucinations have any phenomenal properties. As discussed in the previous chapter, Fish develops this negative disjunctivist approach to defending naive realism. While the claim that total hallucinations lack phenomenal properties, and so have no character, is often met with incredulity, it provides complete immunity to the screening off problem. In addition to providing an account of perfect hallucinations of the sort that feature in the screening off problem, Fish provides a general account of hallucination he intendeds to captures all cases of hallucination. In this chapter, I will primarily be concerned with testing the extensional correctness of his account.

In the first section, I unpack phenomenal disjunctivism and the positive account of hallucination offered by Fish (4.1). In the process of elucidating Fish’s view, I offer friendly amendments to his definition of hallucination to shore up weaknesses. The remainder of the chapter is concerned with two problems for Fish’s account. The first problem involves hallucinations in irrational subjects (4.2). The second problem—which receives extended attention—turns on the HP principle: that hallucinability entails possibility. First, I show that Fish’s view is committed to the HP principle (4.3). Next, I work through several different potential counterexamples to test the HP principle (4.4).

---

95 I’ll assume here that Fish is a mad dog disjunctivist and has a clean escape route from the screening off problem by denying the first premise and is unsusceptible to a revenge screening off problem.
Several different responses and approaches to save the account from counterexamples are evaluated. In the final section, I return to the argument for the HP principle and demonstrate the failure of Fish’s positive account of hallucination (4.5).

4.1 Disjunctivism about Phenomenal Character

Fish proposes and defends a novel *disjunctivist* form of Naive Realism (2008; 2009; 2013b).\(^{96}\) Fish holds that veridical perception involves an irreducible *acquaintance* relation holding between perceivers and mind-independent objects, including some of their properties. While the ‘acquaintance’ terminology comes from Russell, not much turns on Fish’s choice of terminology. He remarks it is the same irreducible perceptual relation that has gone by, ‘appearance’ (Langsam 1997: 36), ‘taking in’ (McDowell 1994: 25), ‘presentation’ (Martín 2004: 38) and ‘awareness’ or ‘receptivity’ (Smith 2002: 43).\(^{97}\) For our purposes, Fish’s ‘acquaintance’ relation can be understood as equivalent to the ‘perceptual awareness’ relation and is consistent with the core theses of naive realism (1.4). Given Fish’s view is consistent with the core thesis of naive realism, I’ll use ‘perceptual awareness’ instead of ‘acquaintance’ in discussing Fish’s view in what follows. The second distinctive feature of Fish’s account is that veridical perceptual experiences enjoy character, while total hallucinations do not. The third feature of Fish’s account is that he denies that the character of veridical perception is narrow. Now, all disjunctivists will likely deny that the character of veridical perception is narrow since character in such cases is

---

\(^{96}\) To be clear, Fish is concerned with naive realism about visual experience. In discussing his views, all claims about perceptual experience are to be understood as restricted to visual experience.

\(^{97}\) Note, the exact ontological category of the *relata* of the perceptual relation may differ across accounts (e.g., for McDowell it is facts, while for others it is objects and their properties, or property clusters). Additionally, there are significant differences between such accounts in cases of hallucination.
determined by standing in perceptual relations to mind-independent entities. But Fish is explicit about his denial. In what follows, I will focus on Fish’s account of hallucination.

4.1.1 Fish’s Account of Hallucination

The following is Fish’s formal definition of a total hallucination:

For all mental events, $e$, in doxastic setting $D$ with cognitive effects $C$ (in its subject), $e$ is a pure hallucination of an $F$, if and only if

- $e$ lacks phenomenal character, and
- There is some possible veridical visual experience of an $F$, $V$, that has a rational subject who is in $D$ and produces $C$, and
- $C$ is nonempty. (2009: 94)

This account is eliminativist about hallucinatory phenomenal character (henceforth, character). A “pure hallucination” is Fish’s technical term for total hallucinations as opposed to mixed, partial hallucinations, e.g., Macbeth’s dagger. Pure hallucinations take place without any background experience of the world and so without any phenomenal properties determined by standing in fundamental relations of perceptual awareness.98

Note, the definition is of a particular token mental event on a particular occasion and is meant to provide an explanation of what it is for a mental event to be indiscriminable from a veridical perception of a certain kind at a particular time.

Next, I will attempt to unpack three important terms used in the formal definition: doxastic setting, veridical visual experience, and cognitive effects. A ‘doxastic setting $D$’ refers to background beliefs, desires, and other mental states of a subject. Fish does not provide a formal definition of ‘veridical’. In characterizing cases of perception as standing

---

98 Unless otherwise specified in the text, ‘hallucination’ will always mean total hallucination (re: pure hallucination).
in perceptual relation to facts he does provide us with a guide to understanding veridicality consistent with the core thesis of naive realism.

On a natural reading of ‘having a veridical experience of an F’, there are cases of having a veridical visual experience of an F that is silent on whether the object is an F. For example, I can have a veridical experience of a dachshund and merely see it as a dog. Thus, on this reading there is a difference between a veridical visual experience of an F and a veridical visual experience of an F as an F. But this is not how the expression is being used by Fish. Following his usage, we will read all instances of ‘a veridical visual experience of an F’ as ‘a veridical experience of an F as an F’; likewise, following Fish we will read ‘hallucination of an F’ as ‘hallucination as of an F’.

It is somewhat puzzling that Fish uses ‘veridical visual experience’ rather than ‘veridical visual perception’ in his formal definition of hallucination since he cannot allow a veridical hallucination (1.2.6). While Fish neither considers nor even comments on cases of veridical hallucination or veridical illusion, they would be problematic if allowed as a veridical visual experience in the definition. If we substitute the disjunction, ‘visual perception or visual hallucination or visual illusion’ in for the expression ‘visual experience’ in the definiens, we end up with a circular definition. Thus, I’ll assume he only means to include possible veridical perceptions in the definition. Putting the last two clarifications together, I will interpret ‘veridical visual experience of an F’ as it appears in the second bullet of Fish’s formal definition to mean ‘veridical visual perception of an F as an F’.

Next, I turn to shedding some light on cognitive effects as used in Fish’s definition. Cognitive effects include but are not limited to beliefs and judgements. Fish fails to provide
a list of everything he intends to include in the category of cognitive effects, but he does
intend to capture all the downstream cognitive outcomes of a perceptual experience that
will help differentiate it from others without appeal to character.\textsuperscript{99}

In following Armstrong’s (1961) approach to providing an account of hallucination
in terms of beliefs, Fish writes, “The phenomenal character of hallucination is simply a
ghost generated by my belief that I am seeing something” (2009: 98). It’s important to
note that although Fish uses the term ‘effect’, a term which is typically associated with
causation, he intends for the relation between perceptual experience and perceptual
judgement to be neutral between a causal explanatory relation and a rational explanatory
relation.\textsuperscript{100} Also, in the formal definition of a pure hallucination it is important that the
cognitive effects, \( C \), on each side of the biconditional match. The second bullet should
read,

\begin{itemize}
  \item There is some possible veridical visual \textit{perceptual experience of an F}
    
    \textit{as an F, V}, that has a rational subject who is in \( D \) and produces \textit{all and}
    
    \textit{only C}.
\end{itemize}

Fish cannot allow \( C \) to only match a subset of the cognitive effects in the possible veridical
case, for it’s only differences in cognitive effects that allows for differentiating various
types of hallucinations. Given Fish denies that hallucinations have character, this is the

\textsuperscript{99} Another example of a cognitive effect that may or may not count as a belief include effects that
can explain unconscious priming, e.g., the influence of subliminal stimuli on behavior.

\textsuperscript{100} According to Fish’s account of cognitive effects, if \( x \) rationalizes but doesn’t cause \( y \) then the
rationalization of \( y \) can count as an effect. A better label might be cognitive upshots.
primary resource for differentiating hallucinatory experience. This will be explored in more detail below.

4.1.2 Could or Would

The second bullet in the formal definition utilizes modality. It is uncontroversial for such a definition to rely on possibilities. However, throughout his discussion it appears that Fish conflates ‘could’ with ‘would’. Here is Fish explaining what the formal definition means,

In plain English, it states that, for a mental event, \( e \), to be a pure hallucination, it must lack phenomenal character yet produce the same cognitive effects that a veridical experience, \( V \), would have produced in a rational subject with the same background beliefs, desires, and other mental states (i.e., in the same ‘doxastic setting’) as the hallucinator. (2009: 94-95)

Contrast the quote above with the formal definition which states that for a mental event, \( e \), to count as a pure hallucination of an \( F \) it lacks phenomenal character but shares the same cognitive effects, \( C \) as some possible veridical experience, \( V \) (re: there could be such a \( V \)).

Given what Fish says, the question is over which of the following versions he intends to be a necessary condition for a mental event, \( e \), to be a pure hallucination of an \( F \):

Could-Version \hspace{1cm} there could be such a \( V \) with the same \( C \), versus,

Would-Version \hspace{1cm} if the mental event, \( e \), were a \( V \) instead, it would have the same \( C \).\(^{102}\)

\(^{101}\) Italics added for emphasis. Fish uses this same expression in defining pure hallucination using ‘would’ several times.

\(^{102}\) Here I’ve omitted doxastic setting and rationality to more clearly draw out the distinction between the versions. I don’t mean for those constraints to somehow not be in play.
Next, I’ll examine the difference between the two versions more closely.

I’ll begin by framing up a case study: Anne is a rational subject. Her background beliefs and desires are of the ordinary variety. She suffers a hallucination of a dog. In order to simplify the comparison in what follows, I’ll hold the following fixed across all the examples: (1) cognitive effects $C$ are non-empty, (2) subjects are rational and (3) hallucinations lack character.

According to the could-version based on the formal definition, to be a pure hallucination of a dog in doxastic setting $D$ with cognitive effects $C$, there only needs to be a possible world (or centered world) where a suitable similar subject in the same doxastic setting $D$ has a veridical perception of a dog with the same cognitive effects $C$. Upon having a pure hallucination of a dog, Anne believes:

(B1) that there is a dog in the room, and

(B2) that there is a chihuahua in the room.

These two beliefs are cognitive effects of the hallucinatory mental event in doxastic setting $D$. It is unobjectionable that there is a possible veridical experience of a dog in a matching doxastic setting as Anne producing the same cognitive effects, B1 and B2. So far, so good for the could-version.

According to the would-version, we evaluate the account by comparing the cognitive effects between the hallucination and the counterfactual scenario involving a veridical experience. First, it is difficult to take what Fish writes literally when he drops the restriction “of an F” in the passage cited above. Without more explicit guidance to identify the relevant veridical experience in the counterfactual scenario, the account fails.
To see this, take the case of Anne hallucinating a dog. Suppose Anne is sitting in her den with her cat. The veridical experience she would have had were she not hallucinating, would have been of her cat in her den. Thus, the cognitive effects would be different. She would not believe B1 and B2. So, we need the ‘of an F’ restriction to specify the kind of veridical experience in the counterfactual scenario that is germane to the account.

But even if we add the kind restriction, ‘of an F’, to the would-version as in the formal definition, fairly straightforward counterexamples remain. Anne has a pure hallucination of a dog. Specifically, she hallucinates a chihuahua. Further, suppose that in Anne’s den where she experiences the hallucination is a saint bernard instead of a cat. Here if we hold fixed the kind ‘is a dog’ in evaluating the counterfactual scenario, we get the following result: If Anne were to have had a veridical experience of a dog, she would not have had the same cognitive effects had she been hallucinating. She believes B1 in both. But she doesn’t not believe B2 in the counterfactual veridical experience. Instead, she believes:

(B3) that there is a saint bernard in the room.

There may be a fix in the offing. In the case of Anne, she hallucinates a chihuahua. The kind ‘chihuahua’ is more specific than the kind ‘dog’. If it was stipulated in the account that the maximally specific kind in the pure hallucination be used to identify the relevant veridical experience in the counterfactual, then these problematic counterexamples are avoided. However, this raises another concern. This only works if hallucinations are always of the same richness or specificity as veridical experience. Given that on Fish’s

view veridical experiences have objects and properties as constituents, while pure hallucinations lack character, it is likely that hallucinations lack the richness or specificity of veridical experience. Accordingly, we will abandon the would-version in favor of the could-version.

4.1.3 Migraine Yellow and Migraine Green

Next, I want to highlight a few preliminary concerns with the could-version as stated in the formal definition and present an amendment to the account. I will explore whether two distinct veridical experiences in the same doxastic setting can have the same cognitive effects. If so, this would make trouble for Fish’s account as we wouldn’t be able to differentiate between the two “intuitively distinct” hallucinations. Bright light is often reported as a trigger for migraine headaches by those that suffer from them. Imagine scientists discover there are two specific kinds of bright light that cause migraines. One such is a particular shade of yellow—call it migraine yellow.¹⁰⁴ The other is a particular shade of green—call it migraine green. Migraine green causes neurological activity associated with low to medium level cranial pain. Migraine yellow causes neurological activity associated with medium to high level cranial pain. Note, migraine yellow and migraine green share a propensity to cause an overlapping range of neurological activity associated with a medium level of pain. Call the cognitive effect of having this type of

¹⁰⁴ Inspired by Kripke’s killer yellow but less morbid. See Lewis (1997: 333). Presumably, on Fish’s view we can’t hallucinate killer yellow, because the cognitive effects are null. Unless killer yellow causes a neurological state that then causes death. If that is how it works, then one can hallucinate killer yellow and it will be equally lethal. One might wonder what deviant mechanism produces the neurological state that the phenomenal experience of killer yellow produces. Fish would say this is exactly what it is to be a hallucination of killer yellow: a deviant cause with the same cognitive effects. To be clear, the original killer yellow kills before any experience of yellow.
neurological activity in the mean of the range with the medium level of pain, M. Now, suppose there is a pure hallucination of a bright light with this same cognitive effect, M. The issue is that we have two possible veridical experiences of a distinct type with the same cognitive effect. So, in such a case there is a pure hallucination of migraine yellow and a pure hallucination of migraine green. And this makes trouble for the could-version.

The problem for Fish’s account in the preceding example is this: his account of hallucination is offered to explain how a veridical perception can be indiscriminable from a pure hallucination. In short, they are indiscriminable just in case when in the same doxastic setting they have the same cognitive effects. But in the preceding example presumably a migraine yellow veridical experience is discriminable from a migraine green experience. Here Fish can point out that presumably they have different beliefs about the color of the bright light in the two cases. Here we can revise the example to include that the neurological activity caused by both migraine yellow and migraine green also disrupts belief formation so that no belief is formed regarding the color of the bright light.\(^\text{105}\) This ensures the same cognitive effects in the two cases.

It appears that Fish intends to deny such cases are possible. He writes, “I suggest that the sameness of cognitive effects actually gives us reason to deny the assumption that the subject really does have two discriminable veridical experiences” (2013b: 62). Fish shifts the burden and holds we need an independent reason for believing there is a discriminable difference in such cases. He argues that the only evidence of phenomenal difference would be differences in downstream effects. But certainly it is possible for there

\(^{105}\) Alternatively, migraine green might always cause false belief/memory that the shade was yellow. See Martin (2013) who also raises this same concern and where one of his counterexamples involves inattentional blindness.
to be two specific cases with two similar yet discriminable veridical experiences which just happen to have the same effects. Since Fish provides an account of hallucinations in terms of their shared cognitive effects with perceptions, he can’t have discriminable perceptions without there also being a difference in cognitive effects.

4.1.4 Cognitive Propensities

There is another option for Fish at this point. Instead of providing an account of hallucinations in terms of sharing their actual cognitive effects with some set of indiscriminable veridical experiences, he might instead build an account out of propensities understood as a range of potential outcomes with a probability distribution associated with various cognitive effects. According to this approach, the same formal definition is utilized but $C$ demarcates cognitive propensity instead of cognitive effects. A hallucination of an $F$ has the same cognitive propensity (outcome profile of potential cognitive effects) as a veridical experience of an $F$ without the character.

This is better. But does the same problem reoccur? Can we imagine a case where we have two veridical experiences with discriminable character but the same cognitive propensity? Remember we are considering two token experiences. It does seem possible that two token veridical experiences with a slight variation in character share the same cognitive propensity. Ultimately, what Fish is after is an account of hallucination that preserves everything it is like to have a veridical perception without the character. Nonetheless, it is counterintuitive to have an account of veridical experience that includes a robust account of character dependent on mind-independent objects and their properties and yet still have to rely on divergences in cognitive effects or cognitive propensities.
(probability distributions of potential cognitive effects) in order to discriminate between them. We would have expected such a naive realist account to provide an account of the discriminability between veridical experiences based on character alone.

4.1.5 Further Qualifications

In continuing to elucidate Fish’s account, notice that there are apparent cognitive effects of veridical experiences that are not available in cases of pure hallucination. First, there may be *de re* judgements made on the basis of the veridical experience that are not possible in cases of hallucination. Fish holds that *de re* judgements must be excluded from the set of cognitive effects (2009: 94, fn. 13). Presumably all such object dependent thoughts based on veridical perception that are not possible in cases of hallucination need to be excluded from the set of potential cognitive effects.\(^{106}\) Second, knowledge of one’s environment is a potential cognitive effect of veridical experience. Suppose I’m having a veridical visual experience of a banana. Given I’m in a standard scenario seeing a banana as it is, there is a formidable tradition in epistemology that holds I also know that there is a banana before me.\(^ {107}\) If this counts as a cognitive effect of a veridical experience, it is not available when hallucinating. Here again, Fish has to also exclude factive mental states as cognitive effects (or potential effects). Here is a governing principle for cognitive effects: If it is not

\(^{106}\) Notice how having even less cognitive effects at his disposal puts even more pressure on providing an account of distinct cognitive propensities for every discriminable kind of veridical perception.\(^ {107}\) Chisholm (1977), Moore (1953), Russell (1948) and Williamson (2000) hold that veridical perception is a form of knowledge in the sense that: if s perceives that p, s knows that p. This is sometimes called the entailment thesis.
metaphysically possible for the cognitive effect to occur in a case of a hallucination, then the cognitive effect is to be excluded.\textsuperscript{108}

4.2 Rationality and Hallucination

I now turn to more substantial problems with Fish’s account of hallucination. Given the formal analysis Fish provides, there are several different types of counterexamples that each take the form of a possible hallucination that doesn’t have a corresponding good case. A corresponding good case of a pure hallucination of an $F$ is a possibility centered on a subject at a particular time having a veridical perceptual experience of an $F$ that satisfies Fish’s formal definition. One type of counterexample includes cases where the cognitive effects involve irrationality. Fish includes a constraint that subjects in his corresponding good cases are rational. He is concerned about correctly categorizing cases of hallucination by cognitive effects.

It may not be immediately clear why Fish includes this constraint. I’ll briefly explain. If he had instead allowed good cases where the subject has strange cognitive effects because they are having a mental breakdown, e.g., a subject forms a belief about dragons on the basis of a veridical visual perception of a toilet flush, then a hallucination of dragons risks getting miscategorized as a hallucination of toilet flushing. After all a veridical perception can produce any cognitive effects you like if we allow the perceiver to be irrational enough. Suppose someone hallucinates a white picket fence with cognitive effects, $C$. If we are allowed to look at any cases of veridical perception, whatever the rationality of the subject then we will have no way of establishing the episode is a case of

\textsuperscript{108} Martin (2013) makes this suggestion and Fish (2013b) endorses it.
hallucinating a white picket fence since if the subject is irrational enough a veridical perception of anything whatsoever could produce $C$.

Fish provides scant guidance on the conception of rationality he puts to work in his definition of hallucination. He writes,

As I intend it [rationality] to be understood, we can say that it is rational for a dog to behave in food directed ways when it perceives food, yet irrational for the dog to behave in the same way when it perceives, say, a mailbox, or an airplane. (2009: 103, fn 24, brackets mine)

This is remarkably underdeveloped. A dog in a cage who sees food might bark and wag its tail. The same dog might also bark and wag its tail when it sees an airplane. So, are we to say the dog is irrational because it behaves in the same way? And what exactly is it for tail wagging to be “food directed”?

Given his definition and examples, Fish employs rationality as a quality of a subject that obtains in virtue of the relationship between sensory experience and judgements. But a subject might be irrational on the basis of her doxastic state alone. For example, take someone who suffers from the paranoid delusion that he is being followed by spiders and that spiders are about to take over the government. The delusion is a part of (or supervenes on) a subject’s doxastic state. This makes trouble for Fish since there will be no good case with a rational subject in that doxastic state. Suppose now our delusional agent hallucinates a spider. This is surely possible. But it looks like it isn’t possible given the kind of theory on the table since there is no rational subject with the same doxastic setting. This has the odd result of not being able to handle a whole class of hallucinations suffered by those also suffering such bouts of irrationality.
Suppose Fish responds by modifying his definition so that corresponding good cases have to be cases in which a veridical visual experience produces the same cognitive effects and those effects are not in themselves irrational (allowing that the doxastic setting may be irrational in a good case). The problem is that when veridical visual experiences of a certain kind occur in specific subjects, they can produce irrational effects. Take the case of Chad who reacts irrationally to his lover, Toby, texting. Nothing in Chad’s background beliefs suggests Toby is having an affair but when he hallucinates Toby texting, he reacts irrationally by thinking Toby is having an affair. This is an irrational cognitive effect. It certainly seems possible that Chad could suffer such a hallucination. But according to the revised definition such a hallucination is not possible because there is no corresponding good case where believing that the subject’s lover is having an affair is a rational cognitive effect. The problem for Fish’s account is that this sort of hallucination case, the sort that produces or accompanies seemingly spontaneous irrational cognitive effects, is exactly the sort of case that we want a positive account of hallucination to handle.

4.3 Hallucinability and Possibility

In the foregoing section, the counterexamples presented to Fish’s account involved there not being corresponding good cases because of the rationality constraint. In the remainder of this chapter, we will explore cases where it is not possible to have a veridical visual experience of an F because it is not possible that something is an F. Take for example, a case of Jekyll and Hyde, where the same person acts, dresses, and appears in two radically different ways such that you don’t recognize them as the same person. Can you have a pure hallucination of a person identical to Jekyll standing next to a person identical to
Hyde? Intuitively, yes. But it is not possible to veridically perceive Jekyll standing next to Hyde because it is not possible for a person to stand next to themself. You can stand next to a twin, clone or doppelganger, but not yourself. One of the differences between hallucinations and veridical visual experiences is that what a hallucination depicts need not be possible. Before exploring a number of such impossible hallucinations, I will first show how Fish’s account commits him to holding that we can only hallucinate about what is possible.

4.3.1 An Argument for Hallucinability entailing Possibility

The following is an argument for the view that hallucinability entails possibility:

(1) Necessarily, if someone has a veridical visual experience of an $F$, then there is some object, $a$, such that $a$ is $F$. (Veridicality Principle)

(2) If someone has a hallucination of an $F$, then it is possible that someone has a veridical visual experience of an $F$. (Hallucinability Principle)

(3) If (necessarily, if someone has a veridical visual experience of an $F$, then there is some object, $a$, such that $a$ is $F), then (if it is possible that someone has a veridical visual experience of an $F$, then it is possible that there is some object, $a$, such that $a$ is $F).$ (Transfer Principle)

(4) Therefore, if someone has a hallucination of an $F$, then it is possible that there is some object, $a$, such that $a$ is $F$. (HP Principle)

\[109\] Perhaps one can stand next to a younger or older self with time travel, but one can’t stand next to themself at the same exact age. Does the hallucination of Jekyll and Hyde include the information they are the same age? Presumably not. This response and others will be explored below.
Notice that premise (1) and (2) are the substance of the argument and follow from Fish’s brand of naive realism and his positive account of hallucination. Premise (3) is an instance of a theorem in modal system K. Next, I’ll examine each of the premises.

4.3.2 The Veridicality Principle

According to Fish’s brand of naive realism, in order for a subject, S, to have a veridical visual experience of an F, V, there is the relevant F-fact serving as a constituent of V. This also serves to elucidate his notion of veridicality. Here are two relevant passages by Fish on this point:

In having facts feature in the presentational character of one’s experience, one is thereby acquainted with objects and their properties; in seeing the fact of the cat’s being black, one sees the cat and its blackness. (2009: 54)

This aspect of the claim enables the naive realist to accommodate the intuition...that it is the very instance of blackness exemplified by the cat that we are aware of when we veridically see a black cat as such. (2009: 59, fn3)

Fish is clear that facts are not abstract but concrete. Facts in their most atomic form have structures, “a’s being F” where a is an object and F is a property. There are also structures, “a’s R-ing b” where R is a relation. It is clear that Fish is committed to:

(V) if someone has a veridical visual experience of an F, then there is some object, a, such that a is F.\(^{110}\)

---

\(^{110}\) When Fish uses the expression *veridical visual experience of an F* (or *hallucination of an F*) he means to be talking about a visual experience of an F as an F. Of course, the ordinary English ‘seeing an F’ need not mean ‘as an F’ -- one can see a stick as a snake. We will read him as meaning ‘as an F’.
Premise 1, the veridicality principle, in the argument above says (V) is a necessary truth. According to naive realism, that perceivers stand in relation to facts when having a veridical experience is not intended to be a contingent matter. If true, it is necessarily so.

It is worth noting that the notion of modality here under discussion is that of *metaphysical* necessity and possibility. Metaphysical possibility is treated as primitive. It is characterized as our most inclusive, maximal notion of possibility. Metaphysical possibility is broader than the notion of physical possibility as the latter is what is consistent with the laws of physics. Logical possibility is generally understood as a notion of possibility in which $P$ is logically possible just in case a contradiction does not follow deductively. Given logical possibility raises questions of what logic and proof theory, it may be logical possibility is the same or more restrictive metaphysical possibility.

4.3.3 The Hallucinability Principle

In moving on to review Premise 2, the hallucinability principle, in the argument above, first note that the notion of possibility invoked in this premise and throughout the argument is metaphysical possibility. Next, the hallucinability principle says its being possible that someone has a veridical visual experience of an $F$ is a necessary condition for there being a hallucination of an $F$. The hallucinability principle follows from Fish’s formal definition of a pure hallucination of an $F$. In order to see how it follows, notice that the following necessary condition is entailed by Fish’s formal definition:

---

Nomological possibility is the general notion for various notions of possibility where $P$ is nomologically possible means $P$ is consistent with a specific set of laws. We can distinguish between various more or less restrictive notions of possibility, e.g., biological possibility is that which is consistent with the set of truths of biology.
For all mental events, $e$, in doxastic setting $D$ with cognitive effects $C$ in its subject, IF $e$ is a pure hallucination of an $F$, THEN there is some possible veridical visual experience of an $F$, $V$, that has a rational subject who is in $D$ and produces $C$

Every pure hallucination of an $F$ is in some doxastic setting and has some cognitive effects in a subject. By Fish’s formal definition, it is a necessary condition for any pure hallucination of an $F$ that it is possible that someone, who is rational, has a veridical visual experience of an $F$ in the same doxastic setting with the same cognitive effects as the hallucinator. The consequent of the hallucinability principle is weaker in the following way: If it is possible that someone, who is rational, has a veridical visual experience of $F$ in the same doxastic setting with the same cognitive effects as the hallucinator, then it is possible that someone has a veridical visual experience of an $F$.

The formal definition of hallucination that Fish provides is for total hallucinations and is intended to rule out partial hallucinations. If you are standing in the veridical perception relation to an object and its properties, then the perceptual experience (what he calls a mental event) of the perceiver will have character. Given his positive account of hallucination is eliminativist about character, he provides his formal definition in terms of a total hallucination to make this point clear. On his account of partial hallucinations, the hallucinated property lacks character and is given a similar treatment in terms of cognitive effects. Any veridical component of a partial hallucinatory experience involves actual objects and instantiated properties and so thereby satisfies the constraint that it is possible

---

112 As mentioned above, according to Fish’s account of veridical perception, subject stands in the fundamental perceptual relation to a fact (state of affairs) that includes an object and its properties. Fish needs a theory for why a subject only sees some of the properties of the object and not all of them. He wheels in a view, called selectionism, in which the perceptual system of the agent selects which properties are detected. In chapter 1, I argued against having facts as the relata at end of the fundamental perceptual awareness relation (1.5.4). I’m ignoring that issue here.
to have a veridical visual experience. Thus, I have dropped the restriction to total hallucination in the argument.

3.4.4 The Transfer Principle

Premise (3) is an instance of a theorem in the weak modal system K. The HP principle deductively follows from the first three premises of argument: the veridicality principle, the hallucinability principle, and the transfer principle. If the conclusion is false, it is unlikely that the transfer principle is the culprit as K is intuitional bedrock as far as modality is concerned. Thus, if the conclusion is false, it is either the veridicality principle or the hallucinability principle that is false (or both).

4.4 Reasons Against Hallucinability Entailing Possibility

In the foregoing, I have shown that Fish’s positive account of hallucination is committed to hallucinability entailing possibility. Specifically, I have argued that Fish is committed to the following principle:

\[(\text{HP}) \quad \text{If someone has a hallucination of an } F, \text{ then it is possible that there is some object, } x, \text{ such that } x \text{ is an } F.\]

Next, I’ll explore examples that serve as case studies and provide reasons against the HP principle.
4.4.1 Moon Jumping Cows

Is it possible that a cow jumps over the moon? That is the same as asking: Is it possible that there is some object, $x$, such that $x$ is a cow jumping over the moon? I suspect here many will hesitate to outright affirm the possibility. It is uncontroversial that it is physically impossible for a cow to jump over the moon. In order for a cow to jump over the moon in its current orbit, a cow would need to jump a height greater than the shortest distance between the earth and the moon which is 225,623 miles.\(^{113}\) There is a related intuition that it is also metaphysically impossible that anything that counts as a cow is capable of jumping such a height. There will be others with the opposing intuition: that it is metaphysically possible that cows are capable of jumping over the moon, perhaps given very different physical laws. For example, one might think it is metaphysically possible that the moon shrinks to the size of a baseball and comes so close to the surface of the moon that cows can easily jump over it.

Whether or not it is possible that there is something that is a cow and jumps over the moon is not uncontroversial after all. On the other hand, it is intuitively clear that one might suffer a hallucination of a cow jumping over the moon. Given the HP principle, suppose you had a hallucination of a cow jumping over the moon and reason as follows:

1. I had a hallucination of a cow jumping over the moon.
2. If someone has a hallucination of a cow jumping over the moon, then it is possible that there is some object, $x$, such that $x$ is a cow jumping over the moon.

---

\(^{113}\) I’m using the progressive verb ‘jumping’, which doesn’t entail success. For example, Sam was building a house. But he died before he built the house. If this is a worry, insert ‘successfully’ before ‘jumping’ as in: X is a cow successfully jumping over the moon.
Therefore,

3. It is possible that there is some object, x, such that x is a cow jumping over the moon.

Would having a hallucination and reasoning in this way resolve controversy over whether or not something is possible? It certainly wouldn’t help to cite your hallucination in making a case for the possibility. Even if it isn’t entirely clear whether the case of a moon jumping cow is a counterexample, that your hallucination wouldn’t be evidence for the possibility counts against the HP principle.

4.4.2 Jekyll and Hyde

Next, consider again the case of hallucinating Jekyll standing next to Hyde. It is uncontroversial that one might suffer a hallucination of a person identical to Jekyll standing next to a person identical to Hyde. Initially, this may appear to be a clear counterexample to the HP principle. However, in defense of the HP principle, one can argue for the possibility of a person identical to Jekyll standing next to a person identical to Hyde via time travel. So, it is possible after all. At least it is possible if the same person is of two different ages and such a possibility is also compatible with the possibility of time travel.\textsuperscript{114} However, it does appear desperate to have to make such qualifications in order to account for the possibility entailed by the HP principle.

There may be another response to such counterexamples to the HP principle. There is a debate about which properties are presented in perceptual experience (Bayne 2009; 

\textsuperscript{114} While time travel is possible, it isn’t entirely clear that humans like us can survive it. In a universe with Einsteinian spacetime, traveling into the future likely would require traveling close to the speed of light.
The two examples just reviewed both involve complex relational properties: being a cow jumping over the moon, being a person identical to Jekyll standing next to a person identical to Hyde. Even if we break these complex properties down into simpler non-relational properties, they still involve natural kind properties: being a cow and being a person.

Both examples also involve singular contents that are about particular individuals. This is very obvious in Jekyll and Hyde, but somewhat more subtle in the case of the moon. We are talking about the moon of Earth not just a moon, i.e., a natural satellite of any planet.

In the case of vision, there is a thin view that holds it is only shape, color, illumination, motion--the so called basic visual properties--that are the constituents of visual experience. No doubt there is further debate about just which properties are to be included as basic on the thin view. For my purposes here we need not wade into those waters. Opposed to the thin view, is the rich view which holds that more than the basic properties--the so-called rich properties--are also constituents of visual experience. The examples of hallucinations of natural kinds and robust features of ordinary objects if taken literally and not reinterpreted in terms of thin properties are only compatible with the rich view. Any view that holds ordinary objects and their properties in our everyday life are presented in our visual experience, hallucinatory or veridical, must reject the thin view.

For certain kinds of intentionalists, we ask what properties are represented in perceptual experience? While the moves available in logical space for defending one side or the other on the thin versus rich perceptual property debate may be different for intentionalists versus naive realists, I don’t think it makes a difference in framing it up at the outset.

Whether personhood is a natural kind has been debated (Kitcher 1979; Wiggins 1976). I don’t mean to be taking a stand. We can rerun the original example with ‘homo sapiens’ in place of ‘person’. Alternatively, personhood even if not natural is a rich property.

There is a further question of just how rich the rich properties of perception are. For example, are moral properties presented in perceptual experience? For my purposes here, most of the examples involve natural kinds and ordinary objects and what appear (perhaps naively) to be recognizable features.
One can defend the HP principle from hallucinations involving cows and persons by rejecting that the rich properties are constituents of any type of visual experiences. On the thin view, there is just an array of thin properties instantiated in the visual field that are presented in visual experience. Cognitive states, e.g., beliefs, may include the rich properties (or concepts thereof) and some are based on perceptual experiences of the basic properties. Strictly speaking, it is not possible to have a hallucinatory visual experience of a cow on the thin view. Instead, when a subject is taken in by their visual experience—whether veridical or hallucinatory—they form a perceptual belief that they are seeing a cow, based on a visual experience of certain basic perceptual properties: shapes, colors, illumination, and motion. So, to adopt the thin view would insulate the HP principle against the class of potential counterexamples—cases of hallucinating impossibilities would presumably be limited to something akin to “contradictory distributions” of shape, color, illumination, and motion. Cases like the two presented, i.e., Jekyll and Hyde and moon jumping cows, would not be viable as I have presented them. But this is not a move available to Fish. First, it goes against a core tenet of his naive realism that we directly perceive ordinary objects such as cars, caterpillars, oranges, and white picket fences and moreover (in the good case) see them as cars, caterpillars, oranges, and white picket fences. Second, Fish (2013a) explicitly argues against the thin view.

In this section, I have presented two potential counterexamples to the HP principle as case studies. Additionally, an argument is made against HP by showing that a hallucination of an F does not count as evidence that it is possible that something is F. I also anticipated a few responses to the potential counterexamples presented. While some

---

118 I’ll explore three examples along these lines below in section 2.5.3.
of the responses appear desperate (i.e., time travel and shrinking moons), the response involving the thin view appears to block the counterexamples from the start. In the next section, I will explore whether the HP principle and Fish’s positive account of hallucination can be defended by adopting the thin view instead of the rich view.

4.4.3 A Thin Defense of Fish’s Account of Hallucination

As already noted, the thin view and naive realism appear to be incompatible. But it is far from obvious whether or not a form of direct realism other than naive realism could defend Fish’s positive account of hallucination from the arguments I have presented against it. Consider a conservative form of direct realism which holds the thin view, conservative direct realism.119 As already noted, it is difficult to imagine a case where one hallucinates an impossibility when all we have to work with are the basic properties. But there may turn out to be such cases that turn on shape and color properties alone. I’ll briefly present three potential counterexamples to the HP principle intended to work against conservative direct realism.

There are two chromatic opponent channels in the human visual system: blue versus yellow and red versus green.120 According to opponent color processing theory, signals about color from cones in the retina are processed in an antagonistic manner such that a response to one color on a channel inhibits the other color on the same channel (Hurvich and Jameson 1956). Thus, reddish green and bluish yellow are called “impossible colors”

119 Scientific realism is sometimes contrasted with naive realism in rejecting the idea that external objects have all the properties we normally attribute to them. I’ve chosen not to use that label as to avoid confusion with other views and commitments with which the label “scientific realism” may be associated.
120 The third opponent channel is black versus white. It detects luminance.
or “forbidden colors” (Billock and Tsou 2010; Helmenstine 2019). I’ll focus on bluish yellow in this example. Bluish yellow is not green. Some shades of green may appear yellowish and other shades of green may appear bluish. But that is not the same as seeing a color that appears both bluish and yellowish in the same place at the same time. Blue and yellow light will cancel each other out and produce yellowish white or white light. Accordingly, there are no bluish yellow objects during natural viewing conditions under any spectrum of light.121 But someone might hallucinate a bluish yellow patch even though it is at least arguably impossible for there to be a bluish yellow object.122 If so, this represents a counterexample to the HP principle based on a basic property so would work against conservative direct realism.

Next, I’ll present two more counterexamples both based on shape. It is easy to imagine two triangles in two-dimensional space coming together to form a square. Specifically, two isosceles right-angled triangles with matching side lengths that come together at their hypotenuses will always form a square. In directly perceiving shape properties in vision, we often are able to quickly detect when an object has equal length sides. In other words, we appear able to directly perceive the property of being equilateral.123 Surely, we can hallucinate equilateral triangles. Likewise, it is hardly controversial that we can hallucinate two equilateral triangles. Further, someone might

---

121 There is research indicating it is possible to experience bluish yellow in a laboratory setting (Billock, Gleason, and Tsou 2001; Crane and Piantanida 1983). Also, Newell (Newall 2021) has recently argued there may be evidence of experiencing bluish yellow from art. An accessible summary of research at the intersection of the cognitive science and the philosophy of color, see Brogaard and Gatzia (2017).

122 Here is not the place to explore whether different laws of physics may make bluish yellow objects possible or whether if lemons looked bluish yellow under normal lighting conditions would make them bluish yellow.

123 The property of being equilateral required for this example is only approximate equilaterality rather than perfect. Additionally, the reliability of detecting even approximate equilateralness turns on viewing conditions and surely has limits, e.g., number of sides.
suffer a hallucination of two equilateral triangles coming together to form a square. But it is not possible for two equilateral triangles to form a square. Two equilateral triangles form a non-square rhombus with a pair of 60° internal angles and a pair of 120° angles, sometimes called a calisson. To be clear, the hallucination of two equilateral triangles coming together to form a square does not include a hallucination of a calisson morphing into a square. If the hallucination as described is possible, it is a clear counterexample to the HP principle.

Let’s look at one last potential counterexample based on shape properties. There are only three shapes that will form regular tessellation. If you want to tile your bathroom floor using a single shaped tile with equal length sides in a uniform pattern such that each of the sides of each tile line up, there are only three tile shapes that will work: an equilateral triangle, a square, and a regular hexagon. So, it’s not possible for there to be a bathroom floor with regular tiling made up of pentagons. And so given HP, it is not possible to have a hallucination of a regular tiled pentagon floor. But intuitively it does seem possible to suffer such a hallucination of a regular tiled pentagon floor. This would represent a third counterexample to the HP principle that would hold against conservative direct realism. It is worth noting that all counterexamples against the HP principle based on basic properties also apply to Fish and naive realism.

Even if turns out that the counterexamples to the HP principle and conservative direct realism do not work, the intuition drawn out in section 2.4.1—*that a hallucination of an F is not evidence for the possibility that something is F*—remains stable on the thin view. Suppose there was a heated controversy over whether it is possible for something to be colored without having any shape, perhaps an infinitesimal extensionless point of color.
It is counterintuitive to think the debate could be settled by discovering that someone hallucinated color without shape.

Given the heavy lifting the notion of rationality does in Fish’s account of hallucination, there is more pressure to develop thin view. The notion of rationality required for Fish’s view is more like a notion of proper functioning of the process between perceptual experiences and the formation of cognitive states. Fish’s positive account of hallucination holds that the possibility of a specific hallucination depends on there being a corresponding good case. For a case to be a corresponding good case it must be a veridical experience \( (V) \) satisfying Fish’s formal definition, which requires that the same doxastic background \( (D) \) and cognitive effects/propensity \( (C) \) are present and there is an appropriately rational connection between the triple \( \{D, V, C\} \). There is a sense in which visual experience is more coarse-grain on the thin view.\(^\text{124}\) It is more course-grain because there are fewer properties presented in sensory experience. Accordingly, an advocate of the thin view will need an account of how coarse-grain perceptual experiences appropriately cause the fine-grain cognitive states, many of which are about rich properties, in a rational subject. For example, one might posit a large repository of conditional prior beliefs that yield a range of rich property beliefs given a specific domain of perceptual experiences.

We can contrast the type of account needed on the thin view with the simplicity of the account given a super-rich view. On the later, cognitive updates about rich properties are rational when they represent the world as presented in experience, absent defeating

\(^{124}\) Seigel (2005, 483) makes this point nicely in terms of the sparse view being “less committal”. She uses an example of plastic fruit versus real fruit in a bowl on the kitchen table. The veridicality conditions for the visual experience are the same. But they are different for the perceptual beliefs.
evidence. Here, the same rich properties are presented in both experience and the newly formed cognitive states. Seeing a Cadillac Eldorado in front of you as a Cadillac Eldorado causes a rational subject to believe that there is a Cadillac Eldorado in front of you. The more liberal the view of the properties presented in experience, the more potential for simplicity in the account of what cognitive effects are appropriately rational for a given sensory experience.

We have seen that the thin view will have to rely more heavily on background beliefs or credences. But none of this will help insulate conservative direct realism from the rationality counterexamples once such realism is combined with Fish’s view of hallucination. Indeed, such counterexamples will well be more plentiful. All we have to do is find cases where the background beliefs are intrinsically irrational or else cases where the background beliefs combined with the thin experience produce irrational cognitive effects. In these types of situations, there won’t be a corresponding good case with a matching veridical experience and where the doxastic background and cognitive effects are found in a *rational* subject. For concreteness, imagine rational updating involves conditionalizing on thin experience. Now imagine an irrational subject who does not conditionalize—they have an experience of $\alpha$ and irrationally arrive at credences that could only rationally be arrived at by conditionalizing on an experience of $\beta$. Fish’s account will issue the verdict that the actual experience was of $\beta$.

Next, I will present one last problem for combining the thin view with Fish’s account of hallucination, one that doesn’t turn on irrationality. I’m going to use a simple model involving the sense of touch. Let’s imagine there are just two pairs of basic properties presented in tactile sensation: (1) the texture of a surface can be felt as rough or
smooth, and (2) the temperature of a surface can be felt as hot or cold. Next, imagine the subject, Sloan, has a prior belief that all rough surfaces are hot such that when Sloan has a veridical tactile experience of a rough object, he forms the belief that the object is rough and hot. There are two possible veridical tactile experiences that are distinct in character but have the same cognitive effects: a veridical tactile experience of rough will produce a belief that the object is rough and hot and veridical tactile experience of rough and hot will also produce a belief that the object is hot and rough. Suppose now that Sloan suffers a hallucination with the cognitive effect of believing that the object is rough and hot. The problem is this: what did Sloan hallucinate? Did Sloan have a hallucination of a rough object or did Sloan have a hallucination of a rough hot object? This is an alternative version of the migraine yellow problem presented above in clarifying Fish’s view.

In defending his view from this kind of problem, Fish holds that any phenomenally distinct veridical experiences will produce different cognitive effects or have distinct cognitive propensities. Perhaps Fish’s best bet is to argue that Sloan has different beliefs about what she experienced in the two cases. While Sloan forms the belief that an object was rough and hot, Sloan believes she felt only roughness in one case and believes she felt both roughness and hotness in the other case. There are two problems with differentiating the cases based on their experiences. First, it seems we can imagine simple creatures that form only beliefs about the world and not about how they experience it. Second, even if a creature could form beliefs about their experience, we can easily imagine a case in which there is no telltale sign to differentiate the experiences by their cognitive effects. The general idea being that there might be a class of perceptual properties where what we believe about objects with those properties is systematically different from what is
presented in experience. In which case, it becomes more difficult for Fish to maintain that whenever there is a difference in character there is a difference in cognitive effects.

This concludes the exploration of a conservative direct realist defense of Fish’s positive account of hallucination. I presented three potential counterexamples to the HP principle, initially viable against both the rich and the thin view. I also showed how two of the main arguments already presented against Fish’s account of hallucination either remain or get worse by combining it with the thin view. Thus, Fish’s positive account of hallucination cannot be adequately defended by adopting the thin view. The main hope for Fish then is to hang tough on the hallucination to possibility principle without having to retreat to the thin view. I admitted that the moon jumping cow and Jekyll and Hyde cases were not totally decisive, i.e., there was some wiggle room to appeal to the moon shrinking and so on. But we have found three more potential counterexamples even within the resources of the thin view. And to finish the discussion, I want to underscore just how plentiful counterexamples will be for the thick view by presenting several more potential counterexamples to the HP principle.

4.4.4 Cat People

John Hawthorne and I have known each other for a long time. We have spent many hours in discussion in close quarters. All this time, I’ve taken us to share species-hood. Suppose

---

125 Research on visual working memory for color indicates that we remember color experience as more closely resembling best examples of color categories (Bae et al. 2015). Study participants are first asked to pick out the best examples of colors like, blue, orange, pink, etc., and then later in the experiment those same study participants will remember various other shades of those colors seen as more closely resembling the best examples they had chosen earlier. Given the human eye can distinguish over a million different colors, working color memory is likely under resourced to recall that many shade variations.
I hallucinate John peeling away fake human skin revealing he had been a cat all along. In other words, I might have a hallucination of a cat identical to John Hawthorne. But it is not possible there is an $x$ such that $x$ is a cat identical to John Hawthorne. (Even if John Hawthorne could grow whiskers and a tail and take to drinking milk from a saucer that would not make him a cat! Being cat shaped is one thing, being a cat is quite another.)

I have already pointed out that Fish excludes *de re* judgements from the class of cognitive effects shared by hallucinations and their corresponding good cases. Cognitive effects that are impossible in cases of hallucination are excluded. But in the example under consideration *de re* judgements are not the issue. In order to defend the HP principle from this counterexample one has to take issue with the property, *being a cat identical to John Hawthorne*, being used as a kind of hallucination. One might argue that properties that are uninstantiable are impermissible as kinds in the definition of a hallucination of an $F$. The reason is that it is only instantiated properties that our visual system is built to detect. Further, it is in this sense that veridical perception plays a primary role in explaining perceptual experience overall. But this defense of the HP principle begs the question at issue.

(We can also sketch variants of this kind of example that do not turn on singular contents. I have a wild hallucination. Cats turn into people. People turn into cats. In my hallucination cats and humans are the same species—cats relate to humans rather as caterpillars are to butterflies—stages in the life of a single species. But it is impossible for humans to be cats, assuming (a) humans are essentially humans and cats are essentially cats and (b) it is impossible for a thing to belong to two species at the same time. Much of what I say below can be adapted to this case.)
We are here investigating whether various hallucinations that have no corresponding good case are possible. To disallow hallucinations of uninstantiable properties is very close to simply restating a version of the HP principle. One might substantiate the restriction by telling a story about the explanatory primacy of veridical perception in terms of what is required for concept acquisition and deployment. But any such maneuvering is at best further support for the HP principle rather than a response to a potential counterexample to the principle.

We can have a veridical visual experience of a cat and we can have a hallucination of a cat. We can also have a veridical visual experience of John Hawthorne, using the same form of locution I’ve rendered this as: a veridical visual experience of a person identical to John Hawthorne. We clearly use vision to recognize specific individual persons. Assuming we are embracing the rich view regarding the properties presented in visual perception, it is difficult to see on what grounds Fish can rule out the property of being John Hawthorne or the property of being a person identical to John Hawthorne as being presented in visual experience? If it is allowed, then it is also possible to hallucinate a person identical to John Hawthorne. Prima facie, we not only imagine and dream about specific people but also suffer hallucinations about specific people. But once it is granted that (1) it is possible to hallucinate a cat and (2) it is possible to hallucinate a person identical to John Hawthorne, how does one then claim it is impossible to hallucinate a cat identical to John Hawthorne? (In thought, we can put together concepts of instantiated

---

126 For example, building on what Robinson (1994) calls the principle of minimal empiricism. The idea being that to grasp concepts about the empirical world requires some form of perceptual acquaintance and not rely solely on internal relations.
properties to generate a concept of an uninstantiated property. Why should this sort of exercise be unavailable when it comes to properties presented in experience?)

Another way of defending the HP principle from this counterexample is to argue that while you can’t hallucinate a cat identical to John Hawthorne, you can hallucinate a large cat tearing away a John Hawthorne mask and costume. This later hallucination is close enough to explain why we have the intuition that the former hallucination is possible. Prima facie, the two hallucinations as described might very well be indiscernible on the basis of introspection alone. If hallucinations have character, one might convincingly argue that there is just one subjective visual character. But Fish has to go another way and differentiate hallucinations on the basis of their cognitive propensity. In the original hallucination case, the subject is likely to believe that John Hawthorne is a cat impersonating a human and in the alternative hallucination case, the subject is likely to believe that there is a cat impersonating John Hawthorne. Intuitively, we can provide a Fishesque account of hallucination based on cognitive propensities and there looks to be two different hallucinatory experiences. In this way, we resist the idea that the alternative hallucination explains away the original counterexample.

4.4.5 Rutgers Philosophy Hydra

Dean Zimmerman and Ted Sider are two philosophers in the Rutgers philosophy department. Like John Hawthorne, I have known them both for a long time and taken us all to share species-hood. I’ve also taken them to be their own person. Suppose I hallucinate Dean and Ted are a two-headed hydra. In other words, I might suffer a hallucination of a two-headed hydra identical to Ted and identical to Dean. But surely it is
impossible that Ted is identical to Dean and so it is impossible that there is a two-headed hydra identical to Ted and identical to Dean. (Note, the hallucination I am describing is one in which I hallucinate that Ted is identical to a two headed hydra. There is another hallucination where I hallucinate instead that Ted is a part of a two headed hydra. That is not my case. It’s not that clear whether Ted could be part of a two headed hydra so it’s not clear that version helps.)

4.4.6 Unicorns

In naming and necessity, Kripke explains it is not possible for there to be an x such that x is a unicorn (1980, 23–24, 156–57). According to Kripke, we first need more information on the internal structure of the mythical species. But that lack of information regarding internal structure doesn’t prevent us from having a hallucination of a unicorn. If the HP principle is correct, Kripke was likely wrong. Further, imagine the following announcement: the next metaphysical mayhem conference will involve philosophers taking hallucinogenic drugs in order to determine what is possible by inducing hallucinations of unicorns, zombies, free will, maximal excellence, etc.

Whether or not Kripke was right about the impossibility of unicorns, there is another defense of HP we have not yet considered. It is what I shall call the composition defense. The main idea being that hallucinations with corresponding good cases can be combined to form hallucinations that don’t have corresponding good cases. There are both veridical and hallucinatory visual experiences of objects that look like unicorn horns, e.g., the tusk of a narwhal. Likewise, there are both veridical and hallucinatory visual
experiences of horses. The composition defense goes on to posit hallucinations of unicorns built out of these other hallucinations which do have corresponding good cases.

Let’s see how this might go. First, according to the compositional account, there are two types of hallucinations: 1) hallucinations with corresponding good cases, henceforth ‘Y-hallucinations’ for hallucinations with a possible matching veridical experience, and 2) hallucinations without a corresponding good case, henceforth, ‘N-hallucinations’ for hallucinations with no possible matching veridical experience. Second, the compositional account adopts Fish’s account of hallucination for Y-hallucinations. Next, the compositional account provides an account of N-hallucinations that employs a notion of composition and relies on the account of Y-hallucinations as a basis. Here is a first pass at such an account:

For all mental events, e, in doxastic setting D with cognitive propensity C,

\( e \) is a total N-hallucination of a \( G \), if and only if

- \( e \) lacks phenomenal character, and
- there is no possible veridical visual perception of a \( G \) as a \( G \), \( V \), that has a rational subject who is in \( D \) and produces \( C \), but
- there are possible total Y-hallucinations with rational subjects in \( D \) each with cognitive propensities that together compose \( C \).

The key move in the compositional account provided above is in the third bullet: that the cognitive propensities of two or more Y-hallucinations together yield the cognitive propensities of an N-hallucination of a \( G \) through composition. But how does such composition of cognitive propensities work?
Given my current purpose of understanding composition in the context of hallucination, the notion of cognitive propensities understood as probability distributions over cognitive effects adds a complication that we can temporarily jettison for heuristic purposes. To simplify the account, let’s replace cognitive propensities with one type of cognitive effect: telltale perceptual judgements. Telltale perceptual judgements are just those beliefs produced by a veridical perceptual experience such that having such beliefs are indicative of that kind of perceptual experience. For example, believing that there is a cat in your field of vision is a telltale perceptual judgement of a veridical visual perception of a cat. We might represent the preceding example for all properties as follows:

For all kinds of visual perceptual experience, $F$, if a subject has a veridical visual perception of an $F$ as an $F$, $V$, then \textit{that there is an $F$ before me} is a telltale perceptual judgement based on $V$.

Presumably, there are many more generalizable telltale perceptual judgments for veridical visual perception of an $F$ as an $F$. But given we now have a working notion in mind, we can plug ‘telltale perceptual judgements’ in for the value of $C$ in the formal definition, replacing ‘cognitive propensity’. The third bullet becomes:

there are possible $Y$-hallucinations with rational subjects in $D$ each with telltale perceptual judgements that together compose $C$.

While we have simplified the account, that only gets us so far. We still don’t have an account of how composition works. As it stands now, composition is a relation between sets of telltale perceptual judgements. The telltale perceptual judgements of an $N$-
hallucination of a $G$ are composed from (built out of) the telltale perceptual judgments of
Y-hallucinations.

In order to examine the relation of composition between telltale perceptual judgments we need to get a fix on just which Y-hallucinations are in play given a specific kind of N-hallucination. Thus far, the definition of N-hallucinations provided has not specified the particular kind of Y-hallucinations relevant for any given kind, $G$. The problem is this: without guidance on the kind of Y-hallucinations, we will get gross misclassification of N-hallucinations. We are attempting to build judgements for a hallucination of a $G$ out of other judgements, but we have not in any way constrained what those other judgments are about.

Intuitively, the Y-hallucinations referenced in the third bullet of the account need to be related in the right way to an N-hallucination of a $G$. Here, we can use the logical structure of properties as our guide. Presumably, any N-hallucination is an N-hallucination of a $G$, where $G$ picks out a complex perceptual property that has simpler perceptual properties as constituents. Further, the simpler perceptual properties are such that they are instantiable; as such, all hallucinations of simple perceptual properties have a corresponding good case. Here is the further revised third bullet of the definition of an N-hallucination:

there is a possible Y-hallucination of an $F_1$ and a possible Y-hallucination of an $F_2$...and a possible Y-hallucination of an $F_n$ for the $n$-many simple properties that are all and only the constituents of the complex property $G$,
each such Y-hallucination with a rational subject in \( D \) and with telltale perceptual judgements that together compose \( C \).

Next, I’ll apply the compositional account to a previous case. Suppose someone hallucinates two equilateral triangles coming together to form a square. The complex property, \textit{being a square made up of two equilateral triangles}, is composed of the property, \textit{being a square} and the property, \textit{being a shape made up of two equilateral triangles}. Since there are veridical perceptions of squares and veridical perceptions of shapes made up of two equilateral triangles, there are corresponding Y-hallucinations with doxastic counterparts. Let’s assume the telltale perceptual judgments of those Y-hallucinations are not limited to but include the following:

(M1) That there is a square before me.

(M2) That there are two equilateral triangles before me.

What is it for M1 and M2 to compose \( C \)? We might look to the logical structure of the property and use that same operation on the distinct beliefs. In the current example, the two constituent properties (\textit{being a square}, \textit{being a shape made up of two equilateral triangles}) are composed through conjunction to form the complex property, \textit{being a square made up of two equilateral triangles}. In the same way, we can conjoin the telltale perceptual judgements of each Y-hallucination. But simply conjoining the beliefs M1 and M2 is not enough to differentiate an N-hallucination of a square made up of two equilateral triangles from an Y-hallucination of a square and two equilateral triangles (an experience as of three objects).
Additionally, suppose that an N-hallucination of a square made up of two equilateral triangles has the telltale perceptual judgement:

(N1) That there is a square made up of two equilateral triangles before me.

This brings me to the critical question: how do M1 and M2 together compose N1? It is natural to think we can break M1 and M2 down into their constituent parts and then use familiar rules of composition. The problem is that relying on basic rules of composition from language and logic, will also yield judgements such as:

(N2) That there is an equilateral triangle made up of two squares before me.

So, there needs to be more specific guidance. How might such guidance be provided?

At this point, one might be tempted to build a framework for representing the hallucinatory visual field for N-hallucinations. The problem is that any such framework will serve as a topology for visual properties. I’m not sure such a proposal will pan out to provide the needed guidance, but if it did, the proposal would be *ad hoc* if were limited to only accounting for N-hallucinations. Further, invoking a visual field representing visual properties appears to be at least functionally equivalent to invoking visual phenomenal properties.

Moreover, in the case of being a unicorn it is very hard to see how the composition solution will work to solve the issue. After all, there could be a horse that grew a horn. But if Kripke is right that would not be a unicorn.

Assuming veridical experiences of unicorns are not possible, Fish might offer a similar line of defense as provided in response to cat people cases. There are veridical
experiences of horses made to look like unicorns. And so there are hallucinations of horses disguised as unicorns. In defending HP against this potential counterexample, one argues that the intuition that a hallucination of a unicorn is possible is explained by the possibility of hallucinating horses disguised as unicorns.

The reply to the surrogate hallucination defense in the unicorn case is similar to the one offered to it by cat people above. There are two different propensities: One in which it is more likely the subject believes they are seeing a unicorn and one in which it is slightly more likely the subject believes they are seeing a horse disguised as a unicorn.

4.4.7 Sun Swallowing Pets

It seems possible for someone to have a hallucination of a dog swallowing the sun. This hallucination is made up of ordinary objects and an action, all of which are perceivable: a dog, the sun, and swallowing. But surely it is impossible for a dog to swallow the sun. The heat of the sun would incinerate the dog before it could swallow it. This is a clear counterexample to the HP principle. And there are many variations on counterexamples of this kind.

---

127 In most medium to large cities, a pony made to look like a unicorn can be hired for fairy tale themed birthday parties.

128 There are many variants involving the sun. I could hallucinate the sun being dragged by a four horsed chariot around the Earth. Indeed, people once believed similar hypotheses. I could hallucinate the sun was made of gorgonzola cheese. And so on.
4.5 Concluding Remarks

The HP principle should be rejected because there are counterexamples. Given the validity of the argument for the HP principle, one or more of the three principles in the argument is false: veridicality, hallucinability, and transfer. The transfer principle is a theorem of K and unlikely the culprit. All forms of naive realism are committed to the veridicality condition. Further, I would argue that a weakened version of the veridicality condition, one in which the properties $F$ are restricted to an acceptable set of sensible properties of objects relative to one’s favored view of perception, is uncontroversial under any standard treatment of veridicality. We can plug any such weakened veridicality condition back into the original argument for the HP principle, thereby, deducing the analogously weakened HP principle—one in which the properties $F$ are restricted to the same set of properties. But even a weakened HP principle that only applies to basic properties is subject to counterexample. Thus, the failure of the HP principle undermines the hallucinability principle and Fish’s positive account of hallucination along with it.
References


