How Things Seem: Arbitrariness, Transparency, and Representation

by

Jessica Anne Heine

B.A. with Honors in Philosophy
Dartmouth College, 2019

Submitted to the Department of Linguistics and Philosophy in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by: Jessica Anne Heine
Department of Linguistics and Philosophy
May 24, 2024

Certified by: Alex Byrne
Professor of Philosophy
Thesis Supervisor

Accepted by: Bradford Skow
Laurence S. Rockefeller Professor of Philosophy
Chair of the Committee on Graduate Students

How Things Seem:
Arbitrariness, Transparency, and Representation

by

Jessica Anne Heine

Submitted to the Department of Linguistics and Philosophy in partial fulfillment of the requirements for the degree of Doctor of Philosophy

ABSTRACT

The principles at the heart of this dissertation are Arbitrariness and Transparency. A representation is arbitrary with respect to content to the extent that the vehicle of representation (brain states, phenomenal experiences, pen marks, soundwaves, etc.) could have represented different content. A representation is transparent if one can be aware of the content of the representation in virtue of hosting it, without being aware of the representation itself.

Chapter 1 argues that visual phenomenal character is fully arbitrary with respect to the objects of perception. That is, for any visually perceptible set of objects and any visual phenomenal character (any ways things perceptually seem) there could be a veridical perception of exactly those objects with that character. This principle is rejected by almost all contemporary theories of perception, yet rarely addressed directly. Many have taken the apparent inconceivability of a certain sort of "shape inversion" — as compared to the more plausible, frequently discussed "color inversion" — as evidence that the spatial characters of our perceptions are uniquely suited to and/or revelatory of the structure of their objects, such that alleged perceptions of those objects that differed radically in spatial character could not be veridical. I argue that these conclusions are unjustified: I claim that the difficulty involved in constructing coherent "shape inversion" scenarios is attributable to the complex relations among visual and tactile shape experiences, as opposed to relations between shape experiences and worldly shape properties.

There is a consensus that endorsing the arbitrariness of perception — as defended in Chapter 1 — necessitates rejecting the transparency of perception with respect to worldly objects. Chapter 2 attacks that consensus. The consensus requires positing a family of properties whose metaphysical status is much more peculiar than is generally appreciated. These "noumenal" properties are allegedly essential to explaining the veridicality of our perceptions, yet no clear explanation is available for how we can learn about them or why we should postulate them. I argue that they do not exist.

Chapter 3 defends an empiricist constraint on understanding language. I argue that the arbitrariness of language prevents anyone — regardless of intelligence, access to data, etc. — from understanding the meaning of words merely by learning how words relate to other words or other arbitrary symbols. While Chapter 1 argues that perception is arbitrary with respect to objects, perception is not arbitrary with respect to perspective-indexed contents. It may be arbitrary how the pen looks insofar as it is arbitrary who is doing the looking, but it is not arbitrary how the pen looks to me or to you. Given what the pen is like and given what I am like, there is only one way for the pen to look to me. I thus argue that one can only understand language by associating at least some linguistic expressions with perceptual representations of parts of the world described by those expressions. If this view is correct, then all knowledge of the world necessarily relies on foundational knowledge about how the world perceptually seems to the knowers.

Thesis Supervisor: Alex Byrne
Title: Professor of Philosophy

Contents

Acknowledgements

Thank you firstly to my wonderful advisor, Alex Byrne, for guiding me through this project. I am deeply indebted to your consistent guidance, extensive feedback, and incredible ability to shepherd my chaotic ideas into actionable plans. I am especially grateful for your generous, intensive support this year as I attempted to defend on a tight timeline.

A profound thank you to Jack Spencer for serving as a constant philosophical mentor throughout my time at MIT. I am immensely grateful for your incredible personal and intellectual generosity. Chatting philosophy with you has been essential to my ability to push forward, especially in times of difficulty, and my coming to feel like a "philosopher".

Thank you also to my other committee members, Roger White and EJ Green, for all of your guidance, support and feedback throughout this project.

I am incredibly grateful for the privilege to spend these years in the community of MIT Philosophy. I am especially thankful for everyone who contributed to ensuring that our community rebounded to full strength after the darkness of the lockdowns. I owe almost all of my ideas to invaluable philosophical engagement with members of this community. For support, feedback, and camaraderie throughout my time at MIT, I especially want to thank David Balcarras, Allison Balin, Kenneth Black, David Builes, Felipe Doria, Kevin Dorst, Bennett Eckert, Jonathan Fiat, Felix-Antoine Gelineau, Selina Guter, Caspar Hare, Michele Odisseas Impagnatiello, Justin Khoo, Abe Mathew, Van McGee, Josh Pearson, Bess Rothman, Kieran Setiya, Brad Skow, Bob Stalnaker, Abi Thwaites, Juan Vargas, Jack Webber, Eliot Watkins, Eliza Wells, Steve Yablo, and Katie Zhou.

Parts of this work have benefited from valuable feedback at a variety of venues, formal and informal, within MIT Philosophy, as well as the 2023 Eastern and Central APAs, the Metaphysical Mayhem Workshop at Rutgers, and several anonymous reviewers.

Thank you to MIT Women's Water Polo, which became a wildly unexpected, incredibly important part of my life at MIT, providing much needed breaks from philosophizing. Thanks especially to Amanda Jackson, Soumya Kannan, Sasha Konavalenko, Pip Knight, Corissa Lee, Sophie Reynolds, and Abby Smith.

I am grateful to Dartmouth Philosophy, where I was first introduced to this incredible world. Thank you especially to John Kulvicki, David Plunkett, Sam Levey, and Amie Thomasson for getting me started on this path.

Thanks you my sisters Meghan and Jennifer for their support. Thank you to Jen for all of the meticulous proofreading over the years! (Remaining typos are probably your fault.)

Finally, I owe most thanks to my parents for all of the love and support that have made this possible. Thank you for empowering me to follow my passions wherever they lead.

Introduction

Many things have a certain species of property: they seem certain ways to certain possible and actual perceivers. My project is about taking these properties seriously.

This red mug looks differently to me from different perspectives. Presumably, the mug looks still differently to animals. The mug may look ever more differently to Martians. For any way of perceptually seeming, the mug would seem that way in some possible circumstance. According to prevailing wisdom, the vast majority of ways the mug may seem fail to reveal what the mug is really like. In Chapter 1, I reject this view. I argue that any way of seeming could — in the right circumstances to the right perceiver — truly reveal what the mug is like.

Most presume that if any way of perceptually seeming is potentially as good as any other, then perception must be incapable of revealing what the world is really like. In Chapter 2, I argue that this widespread view requires presuming that "what the world is really like" features peculiar properties that do not constrain how anything seems, but merely determine which seemings are veridical. I deny the existence of such properties. I advocate for restraining our realism to the so-called world of appearance.

In Chapter 3, I argue that understanding language — and thus understanding anything learned via language — requires associating at least some linguistic expressions with perceptual representations of parts of the world described by those expressions. Understanding how language relates to other language or other arbitrary representations can never suffice for understanding how language describes the world. Knowledge of how things seem must be our foundation for all representation of the world.

Chapter 1: Against Character Constraints

ABSTRACT: This chapter defends the following principle: For any visually perceptible set of objects and any visual phenomenal character, there could be a veridical perception of exactly those objects with that character. This principle is rejected by almost all contemporary theories of perception, yet rarely addressed directly. Many have taken the apparent inconceivability of a certain sort of "shape inversion" — as compared to the more plausible, frequently discussed "color inversion" — as evidence that the spatial characters of our perceptions are uniquely suited to and/or revelatory of the structure of their objects, such that alleged perceptions of those objects that differed radically in spatial character could not be veridical. I argue that these conclusions are unjustified: I claim that the difficulty involved in constructing coherent "shape inversion" scenarios is attributable to the complex relations among visual and tactile shape experiences, as opposed to relations between shape experiences and worldly shape properties.

1. Introduction

I currently see a silver laptop, a red stapler, a blue mug, and some other desk items. There is something it is like for me to see these objects that differs from what it is like for me to see the leafy trees outside. Could some creature have an experience with the same phenomenal character as this, yet see the trees? Most would accept such a possibility: one can see what are in fact leafy green trees, even though these objects appear like a silver laptop, red stapler, and blue mug. However, it will be insisted that such an experience would necessarily be a case of dramatic misperception. After all, the trees are not metallic, red, rectangular, etc. Perhaps this reasoning is sound with respect to humans. However, this chapter argues that such a perception, if had by the right perceiver, need not be falsidical. For some possible creature, a perception with the same phenomenal character (henceforth "character") as my current experience *would* reflect what the trees are really like and would not represent the trees as being any way they are not. Put more generally:

> *Full Permissibility (FP):* For any visually perceptible set of objects (in a certain spatial arrangement with a given set of properties[1]) and any visual phenomenal character, there could be a veridical perception of exactly those objects with that character.

FP is restricted to vision purely for ease of exposition. Although FP is rejected by most contemporary theories of perception,[2] Papineau is a notable potential sympathizer. He argues that "conscious sensory properties . . . represent worldly facts . . . only in virtue of further contingent facts about the way they are embedded in the wider world" (2021, 5). If characters are only related to their objects contingently, then perhaps an experience phenomenally like my experience of my desk could veridically represent leafy trees.[3] Another instance of permissibility-friendly reasoning is found in Chalmers' view of virtual reality, according to which someone with an experience phenomenally like mine could be veridically perceiving something radically different than what I perceive, such as a virtual desk instead of a physical desk — although, perhaps not leafy trees instead of a physical desk, as FP requires (2022; 2016).

[1] That is, if we take a set of objects that can be seen together, and we take a visual phenomenal character, there could be some creature who veridically perceives those objects veridically via that character, even holding fixed the objects' spatial relations and properties (beyond those of the sort, "is seen by S.") 'Objects' can be understood inclusively, such that parts of objects count as objects.
[2] Examples of views that reject FP include Pautz 2021; Levine 2018; Mendelovici 2013; Thompson 2010; Siegel 2006; Chalmers 2004.
[3] Notably, while not taking a stand on FP, Papineau *does* place constraints on which characters can represent which properties: "Square and circular shapes in the world themselves have a structure that an adequate system of symbols for representing them needs to match." (2021, 111)

This chapter argues against perceiver-independent constraints on the character of veridical perception. Philosophers of perception more frequently discuss the determination (or lack thereof) of character by perceptual content. You might think that the question answered by FP — which characters are suitable for veridical representation of which bits of the world? — only arises if character depends on (or is determined by) the subject rather than the perceived world itself. If there is nothing more to character than how the world is perceived to be or which scene a perceiver is related to (as some Naïve Realists hold), perhaps you can reject FP on this basis alone, without any principle for which characters are suitable for veridical perceptions of which objects. In section 2, I argue that no notion of dependence can play such a role. In some sense, everyone must accept that character depends on both perceived objects and something further (plausibly, perspective). Everyone accepts some cases in which objects are held fixed, yet characters of veridical perceptions diverge. Once some such divergence is accepted, and an accompanying explanation is provided, the question arises as to the extent of such divergence — a question that is independent of Naïve Realist and Intentionalist theses about character and content. In sections 4 and 5, I argue that the only justifiable principle rejects all character constraints on veridical perception.

2. Character Determination and Divergence

Theories of perception are frequently sorted to the extent that they attribute the "determination" of the character of veridical perception to features of the mind-independent perceived world as opposed to features of the perceiving subject. For instance, Logue categorizes theorists on a spectrum ranging from those claiming that the character

of veridical experience is "*entirely* determined by the features of the subject" to those who claim that the character of veridical experience is "entirely determined by the properties one perceives of the *mind-independent objects* one perceives" (2012, 214, 216).[4] Yet, focusing on the determination of character can be misleading, given the indisputable role of both subject and object in determining character. Take my veridical perception of my stapler. The stapler causes changes to me via my eyes, causing me to enter a certain brain state. In this state, I am aware of my experience's character — that is, of what it is like to see the stapler. Every believer in the perception of mind-independent things will accept ways that the objects of my perception could have differed and ways that the subject could have differed such that the experience would have differed character-wise. For instance, if the stapler were blue instead of red, curved instead of boxy, further to the left, etc., the character would have differed. My visual tracking of these aspects of the stapler is part of why I count as seeing the stapler veridically. For another perceiver of this stapler in another setting, other properties might be relevant. Yet, any perceiver who genuinely sees the stapler as it is must — at a bare minimum — track *some* aspect of the stapler. Someone whose visual experiences were not correlated with or counterfactually sensitive to any aspect of the objects affecting their eyes would not count as veridically seeing — or, perhaps, seeing at all. In this sense, everyone posits some dependence of character on perceived objects. Conversely, everyone accepts the empirical fact that if the subject of this experience had differed in various ways — such as by having different brain structures, perceptual apparatus, location, etc. — the character of the experience would be different. Yet, *there is* substantial disagreement about how much, and in which ways, a

[4] Pautz's characterization of Naïve Realism versus Representationalism echoes this framing (2023).

given experience could have differed without impacting veridicality. More generally, there is disagreement over the extent to which fixing the features of perceived objects fixes character, provided the scope is limited to veridical perception. I argue against any such constraints.

The polar opposite of FP denies *any* variation between characters of veridical perceptions of a given set of objects. It is difficult to imagine anyone endorsing such a principle, given that the same objects apparently appear differently from different spatial perspectives. For instance, imagine S1 and S2 looking at the same three sides of the same white cube floating in a black void, with S1 being closer to the cube than S2. Despite seeing the same object, their experiences differ in character because their perspectives differ. Here is a minimal principle for character divergence:

> *Minimal Permissibility:* For some visually perceptible sets of objects, veridical visual perceptions of exactly those objects can differ in phenomenal character.

Accepting veridical perception of mind-independent worldly objects while denying *Minimal Permissibility* requires accepting that each set of objects can only be veridically seen from a singular point in space. Given the arbitrariness of any such spatial perspective being the "one true perspective" on any given objects, such a position is tough to defend. Everyone who posits veridical perception of mind-independent objects will accept that some instances of perception fit this "character divergence" template:

> 1) At time t, S1 and S2 both veridically perceive only W.

2) S1's perceptual experience of W at t differs in character from S2's perceptual experience of W at t.

In such cases, one cannot appeal to a difference in the world to explain the divergence given that both perceptions feature same objects. Thus, the notion that *what the perceived bits of the world are like* can entirely explain the character of veridical perception is untenable. Further explanation is needed. Here are two available explanation types for character divergence:

Anti-Intentionalist Explanations: Rejecting the Intentionalism thesis that character supervenes on content, one may claim that properties appear differently character-wise from different perspectives. For instance, S1 and S2 veridically perceive the same properties of the cube, such as the squareness of its surfaces, yet those properties appear differently from different perspectives. S1's and S2's characters differ shape-wise because squares look different from different points in space. Whether or not any subject S veridically perceives a surface's shape depends not only on the character of S's experience and the surface's shape but also on S's perspective. To know whether S veridically perceives the shape, we must know not merely whether the character is appropriate for seeing squareness but whether it is appropriate for seeing squareness from S's perspective relative to the surface in question. The same can be said for size and location.

Intentionalism Compatible Explanations: Alternatively, one could claim that owing to their different perspectives on W, the subjects perceive different properties of W, leading to different characters. (I employ 'perspective' liberally such that any aspect of S1 that allows S1 to perceive features of W that other perceivers of W do not perceive may be considered

a feature of S1's "perspective" on W.) If veridical perceptions feature the same objects and properties, then they are identical in character (because content entirely determines character in veridical perception). However, S1 and S2 do not represent the same properties. Neither misperceives the cube; rather they each perceive a different, incomplete set of the cube's properties.

To fill out this explanation, one might appeal to perceiver-relative properties. Although S1 and S2 perceive all of the same inherent, subject-independent properties of the cube, perhaps they also perceive properties of their own relations to the cube. For instance, it might perceptually seem to S1 that S1 is located slightly above and to the left of the cube. It doesn't perceptually seem to S2 that S1 is related to the cube in any particular way. Such relational properties constitute a distinction in the content of S1's and S2's perceptions, explaining the difference in their character. Campbell suggests this sort of explanation by describing perceiver-relative properties as part of the "constitution" of character:

> [T]he 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* (2002, 116, emphasis added).

Another Intentionalism compatible explanation appeals to properties that are not subject-dependent yet are perspective-centric in that a subject perceives them because of properties specific to that subject. Perhaps S1, but not S2, perceives the way in which the cube is spatially arranged relative to the point in space that S1 (or S1's eyes) occupies. Explanations involving such perspective-centric properties are standard for cross-modal cases of character divergence. If S1 sees W while S2 feels W, most would agree that S1's

character differs from S2's character in so far as S1 is aware of W's visual properties while S2 is aware of W's tactile properties.

It may be helpful to see how a non-Intentionalist and an Intentionalist offer parallel explanations for a given instance of character divergence. Peacocke, a non-Intentionalist, offers a case of a subject looking at two identically sized trees, one closer and one further away (1983). Just as the properties of the cube seem insufficient for determining the character of S1's and S2's experiences, the properties of Peacocke's trees seem insufficient for determining the character of his subject's experience. Although the trees are identical in size and although the subject seems to be seeing veridically, "there is a sense in which the nearer tree occupies more of [S's] visual field than the more distant tree" (1983, 83). Peacocke frames this puzzle as "the problem of the additional characterization", arguing that characterizations of the content of veridical experiences are insufficient to explain the character, and thus additional characterizations, beyond the perceptual content, are required (1983, 83). Byrne, an Intentionalist, offers an account of this case that appeals to perceiver-relative properties to incorporate the asymmetry between the trees directly into the content description: "It visually appears to the subject that he's facing two similar-sized trees, one further away than the other" (222). Both accept that features of the subject's perspective are required to fully account for the character; the disagreement lies only in whether these perspectival features belong to the content.

This section has shown that everyone who accepts *Minimal Permissibility* — which is tough to deny — must offer some explanation of character divergence. This work is neutral between available explanations. Whichever explanation you prefer, I aim to

convince you that you ought to endorse analogous explanations of analogous, yet more radical instances of character divergence among creatures reliably tracking the same bits of the world. I have emphasized that everyone must accept that a single set of objects can be veridically perceived via perceptions that differ in character, even within perceptual modalities. Facts about what the perceived bit of the world is like cannot entirely determine the character of veridical perception; features of the perceiving subject must play some role — whether because properties appear differently via different perspectives and/or because different properties are perceptible from different perspectives.

3. Perspectival Facts

Most would accept a version of FP that omitted 'veridical'. That is, most accept that any character is consistent with the *perception* of any objects, so long as misperception counts as perception. For instance, perhaps I could see the objects on my desk via pink-elephant-esque character. I would simply fail to see these objects as they truly are. For simplicity's sake, in what follows, I will take for granted this weaker claim about the relationship between perception and character.[5]

Stipulate that S perceives objects W via character C. As argued above, knowing all there is to know about W and C is (at least sometimes) insufficient for determining whether S veridically perceives W. Everyone who accepts *Minimal Permissibility* will accept that

[5] You can avoid this assumption — rejected by some, such as Dretske 1969 and Montague 2013 — by replacing cases of "seeing" with "quasi-seeing", such that a quasi-seer of object o visually tracks properties of o on the basis of direct visual input from o. My arguments that some subject S veridically rather than falsidically sees o can also generate the conclusion that S *sees* rather than merely quasi-sees o.

there could be more than one character that could be the character of a veridical perception of W. For instance, in our initial example of S1 and S2 seeing the same white cube, knowing all there is to know about the cube and the character in question is insufficient for knowing whether a perception is veridical. If you are omniscient about the cube and you know that some perception of the cube had C1, you still cannot know whether or not the perception was veridical without learning more about the subject. If the subject was S1, but not if the subject was S2, then the perception is veridical. Everyone must accept that some additional characterization, beyond all facts about the objects and character, may be required to ascertain the veridicality of the perception.

Different explanations attribute different roles to these additional facts. According to anti-Intentionalist explanations, we need to learn facts about S's perspective on W because W seems different character-wise from different perspectives. According to Intentionalism-compatible explanations, we need to learn facts about S's perspective on W because different properties of W are perceptible from different perspectives. Any of these explanations of character divergence could in principle be used to claim that any perceiver of W veridically perceives W. Whichever aspects of S make it the case that S experiences C when perceiving W might occupy the role of S's perspectival facts. Via anti-Intentionalist explanations, we might say that from this particular perspective of S, W-like things bring about C-ish character when veridically perceived, just as we say that from the perspective of someone closer to one of two identical trees, one tree occupies more of the visual field. Via Intentionalism-compatible explanations, we might similarly identify the properties of W that are relevant to S's perspective, describe these properties as part of the content of S's

experience, and claim that C is the one and only character of veridical perceptions of such content.

Of course, endorsing an explanation of character divergence in some cases in no way entails that this explanation need apply to all potential cases of character divergence. However, nothing about the structure or content of these explanations entails that they ought to be offered to justify some particular level of character divergence. One could endorse any of these explanations (and thus endorse or reject Intentionalism) alongside any principle from *Minimal Permissibility* to FP. Thus, to accept *Minimal Permissibility* and reject FP, one needs some further principle for deciding how much character divergence is possible and which perspectives on the world can facilitate veridical perceptions. If one denies that a perception is veridical, this denial thus cannot be justified merely by pointing out that the perception differs character-wise from a selected veridical perception (i.e. "W is like *this*, so W can't be like *that*."). Since everyone accepts such divergence in some cases, character divergence alone cannot show that both perceptions are not veridical. Nor can Intentionalism provide independent reason to reject a perception, given available Intentionalism-compatible explanations. One might assert that which characters are suitable for veridical perceptions of which objects is simply an evident, brute fact. Or, perhaps some kind of suitability criterion might be offered to systematically identify suitable characters for given objects. The rest of this chapter argues that no restrictive suitability principle is justified.

4. Character Divergence Among Simple Perceivers

Imagine two species of simple, mostly immobile perceivers: Normals and Weirds. Their only perceptual capacities are functionally equivalent visual systems that track the same few properties of a few sorts of objects in certain circumstances. Perhaps these objects/properties signal when they should open their mouths to eat passing prey. Both perceivers have white′ visual experiences when seeing nothing. (I follow the convention of using the "prime" (′), introduced by Peacocke 1983 to indicate qualities of perceptual experience while remaining neutral on the relationship that experience has to the outer world. For instance, if an experience is red′ or triangular′, it is phenomenally like the paradigmatic human visual experience of redness or triangles.)[6] One important object for these creatures is a black right triangular prism, with triangular bases with 3-foot edges. When the triangular base of this block is 10 feet away from our perceivers' eyes, centered and perpendicular to their gaze, with the bottom edge parallel to the ground, the creatures see the block and visually track any width changes. If we replaced this special "triangular block" with something visually indistinguishable, such that our creatures' eyes received the same input they receive from the block, the creatures would have the same visual experience as when seeing the block. Similar stories can be told about other objects, such as black cylinders, etc. Normals see these blocks via the characters of a normal human's perception of the relevant block floating in a white abyss. Weirds see the triangular block via monochromatic gray′ character, like a human's character when seeing a gray abyss. As

[6] Some might reject that experiences of different objects can have the exact same character. Still, it is undeniable that the characters of different experiences can be indiscernibly similar. We can say that such experiences share character C, even if finer distinctions are available.

the triangle's width approaches 0, the Weird's character approaches white′. As the triangle's width increases, the Weird's experience approaches black′, until, once the width of the triangle is sufficiently long that Normals could not see the triangle's edges, the Weird's experience is black′. Weirds see the other notable objects similarly through monochromatic′ experiences, such as by tracking the small to large cylinders via blues′.

First, to accept that Normals veridically see these blocks, consider fictional seers with full-body eye-like surfaces, in contrast with our two eyes, who can see clearly in any level of light and differentiate colors (or color-like properties) across a vast range of the electromagnetic spectrum, in contrast to our small window of visibility. Just as we recognize the limitations and arbitrariness of Normals' visual system, these "superseers" will recognize the limitations and arbitrariness of human visual systems. It is difficult to imagine a non-arbitrary, non-anthropocentric standard for veridical perception that would rule human vision sufficient and Normals' vision insufficient.

As Normals and Weirds are visually sensitive to the same objects and properties, their visual systems provide equivalent functional benefits. You might think that, *at least*, only Normals can see the angles and sides of the block. Yet, Weirds are equally visually sensitive to such parts insofar as any difference in size or angle would prevent the Weirds, along with the Normals, from seeing the block — even if the Weirds fail to see the sides and angles *as* sides and angles. For any useful action that a Normal can take on the basis of their perception (i.e. opening mouth to eat, tilting towards light, etc.) a Weird can take the corresponding action in response to their corresponding perception. One might presume that Normal perception is superior given how it might more readily evolve into a more

useful system, such as by allowing for perception of different sorts of triangles or integrating tactile perception. Yet, so too could the Weird system evolve. (Section 5 considers more sophisticated creatures.)

Normals and Weirds respectively develop languages that purportedly refer to perceived properties, such that if they lived together, they'd develop a smooth Normal-Weird apparent translation scheme. Still, many will claim that while Weirds' perception is useful, Weirds simply do not see the blocks as they really are. For instance, their perceptions of triangular blocks are all light grayish′, while the blocks are black. These experiences are shapeless′, while the blocks are triangular. Perhaps, Normals veridically perceive the block as triangular and black, while Weirds do not veridically represent any property of the block. Given the equivalence in functionality, this claim will, in some form, deny the suitability of the character of Weird perceptions. A theorist with this view who adopts anti-Intentionalist explanations of other instances of character divergence might argue that grayness′ and blueness′, for example, are unsuitable ways for triangularity and circularity to appear. A theorist who adopts Intentionalism-compatible explanations elsewhere would need to argue that there are no properties of the block that Weirds veridically see via grayness′ and blueness′. While Normals veridically see triangularity via the triangularity′ of their perceptions, Weirds do not veridically see any property via their grayness′. Yet, this theorist must embrace a fairly permissive view of perceptible properties, given that they agree that S1 perceives properties of that white cube in the black void that S2 does not perceive, such as the angle at which the cube is tilted with respect to S1 or S1's location. In the Weirds' case, there *is* some identifiable property g (perhaps triangularity, triangularity of a certain kind of block, some other coextensive

Weird property) such that if W has g and a Weird sees W, their experience will be grayish′. Why is this insufficient for veridically perceiving g? Is g not the sort of property that can be perceived? Is grayishness′ unsuitable to g?

Either account ultimately relies on claims regarding the suitability of various qualities of perceptual experiences to various properties of worldly objects. While different creatures may be equally reliable in tracking objects' worldly properties via their phenomenal experiences, only the characters of some such creatures' experiences' characters reveal/reflect/represent/etc. what the world is really like. The difficulty is that making non-trivial sense of these claims apparently requires some substantive notion of what a worldly property is like that is entirely independent of our substantive notions of what it is like for us to perceive that property. This sort of conceptual confusion involved in distinguishing what it is like to perceive a property from what that property is like is sometimes invoked in arguments against the possibility of character varying independently of content. In "Intentionalism Defended", Byrne imagines a study in which lay folk are taught the language of phenomenal character and then questioned about the characters of their experiences. Subjects are shown three separate chips separately. The first two are red and the last is blue. "All can agree that, insofar as "what it's like" to undergo the experiences is concerned, seeing the first chip and seeing the second have something in common that seeing the third lacks" (2001, 206). They are taught to record this difference by saying that the first two experiences have the R-character, while the last has the B-character. Byrne points out that a subject clearly misunderstands this new concept who claims, while looking at a blue chip, that his experience has suddenly shifted from the B-character to the R-character even though the chip continues to look blue to him. The subject cannot assess

the character of his experiences of the chip independently of his assessment of how the chip perceptually seems to him to be.

The confusion of Byrne's imagined subject, I suggest, occurs more subtly on a larger scale when philosophers assume they can independently assess a phenomenal property and a worldly property to decide whether or not that phenomenal property is suited to that worldly property. While the confused study participant is only considering one property and one experience, we have endlessly many interrelated properties and experiences. However, all of our knowledge of worldly properties ultimately depends upon perceptions of the world. For instance, we might begin with visual representations of triangles, and then derive principled information about all possible triangular′ representations (i.e. the angles must be 180°), yet this alone tells us nothing about the suitability of triangular′ or non-triangular′ perceptions of triangular things. Or, we might analyze whether our visual perception of the block jibes with our tactile perceptions or more precise visual perceptions, such as those that include measuring devices. Yet, it is rather mysterious how we might assess whether *all* of our perceptions somehow fit the world that we have learned about through those perceptions.

If you know that Normals, but not Weirds, veridically see the blocks, could a Normal and/or Weird discern this information? Normals have no more reason than Weirds for believing that they veridically perceive the world. There is no unique internal feature of Normals' experiences, as compared with Weirds' experiences, that justifies belief in the veridicality of their perceptions. How would they ask whether their reliable perceptions reflected what the world was really like? Such a question is about how one's experiences

relate to how things are. Yet, both the Normals' and Weirds' concepts of worldly properties were christened, taught, and learned in conjunction with their own perceptions. Weirds' concepts of worldly properties, such as g-ness, were developed and learned in conjunction with the properties of Weirds' corresponding grayish′ perceptions. Non-philosophical Weirds likely lack any concept of grayishness′; they will only think in terms of the posited worldly property of g-ness. If Weirds fail to veridically perceive the world, then at least one of the following must be true:

A) Grayishness′ is unsuitable for perception of g-ness.

B) g-ness is not visually perceptible.

C) The triangular blocks are not g.

Supposing A were true, then, as 'g' was coined by Weirds with grayish′ experiences talking about objects that like *that*, contradictorily, 'g' would not refer to g. Weirds would have no reason to endorse B or C beyond an abstract skepticism that their perceptual experiences are somehow unsuited to the world they encounter. (They might also wonder whether their perceptions fail to reliably correlate with worldly properties, but we know such hypotheses to be false.) Normals would be equally well justified as Weirds in supposing they weren't veridically perceiving the world.

One might allege that the Weirds are simply in an unfortunate epistemic state. Yet, the Normals and the Weirds have functionally equivalent perceptual systems. The only difference between them arises from what it is like for them to have various perceptions. Whichever evolutionary pressures give rise to Normals could equally well give way to Weirds. Furthermore, if Weirds don't veridically perceive, then, as there are infinitely many

potential varieties of equally evolutionarily fit "Weird" perceivers with similarly "weird" perceptual characters, presumably even Normals cannot justifiably believe in their veridical perception. Some may accept this conclusion while insisting that humans *can* ascertain the general veridicality of their perceptions. Perhaps the complexity of human perception rules out the possibility of radically divergent, functionally equivalent perceivers. Section 5 considers this view.

5. Character Divergence Among Complex Perceivers

It is a common view that there is no unique suitability between hues′ and color properties, such that aliens who experience green things reddishly′ might yet perceive those green things veridically. Extensive debate has focused on the potential for creatures with behaviorally undetectable inverted spectra, such that they experience green things reddishly′, red things greenishly′, etc. yet behave indistinguishably from normal color seers. People have often reasoned that if such creatures are possible, we have no reason to rationally conclude that our own color experiences are more revelatory of the nature of reality than the invert's experiences.

Many have taken the apparent inconceivability of a similarly radical inversion with respect to shape as evidence that the spatial characters of our perceptions are uniquely suited to and/or revelatory of the structure of their objects, such that perceptions of those objects that differed radically shape′-wise could not be veridical.[7] The sort of shape

[7] Recent versions of this sort of claim are found in Papineau 2021; Levine 2018; Logue 2012; Thompson 2010.

inversion typically considered involves a systematic visual character inversion of two shape′ properties, such as squareness′ and circularity′, without any other perceptual change. Given the complex relationships among shapes, it is extremely difficult to imagine this sort of inversion in a well-functioning individual. From this point, much has been concluded regarding the relationship of spatial character to worldly shape properties. For instance, according to Logue, while color phenomenology is mostly explained by contingent facts about our visual systems, "when it comes to the phenomenology of *shape* experience, it's hard to resist the conclusion that the shapes *themselves* are doing most of the work in determining phenomenal character" (2012, 216). I suggest that the alleged impossibility of the favored form of spatial inversion only provides information about the relation that various phenomenal qualities have to one another — such as visual rectangularity′ to visual triangularity′, etc., as well as visual spatial qualities to tactile spatial qualities — as opposed to the relation of characters to various worldly properties.[8]

Here is an imperfect analogy: Try to imagine a language just like English in every way except that the meanings of the following character strings are inverted: 'That cat is white', 'Are you an American citizen or a German citizen?' All other linguistic meanings are held constant, including the oral versions of these sentences as well closely related character strings such as 'Are you an American citizen or a British citizen?', 'That cat is not white', 'This cat is white', etc. We could certainly adopt such a language effectively in some official sense, perhaps as part of a code or game. Yet, it is difficult to imagine that humans could have such a language as their native tongue without there being *some sense* in which

[8] I'll ignore "Molyneux's Question", as my argument is consistent with any answer (1688).

these special strings retain their English meanings or at least bear special semantic relationships to those meanings. It is not merely that such a language would be strange and impractical. Rather, if the speakers are human-like, it seems that these strings *must* bear some special semantic relation to their English meanings, given the meanings of their parts and general compositional and phonetic patterns. Perhaps the sense in which 'This cat is white' would retain its English meaning would be similar to the sense in which "awesomer" means *more awesome* despite the former's lack of "official" status. If we consider longer, more complex strings — or perhaps oral expressions, which take longer to process — particular inversions of specific linguistic representations that leave the semantic properties of all other linguistic representations undisturbed become even more unfathomable.

This result is explained by the complexity of the relations between and patterns among our linguistic representations, rather than some feature of the relations between representations and their worldly objects. To have a language containing an infinite variety of meaningful linguistic representations, how each representation attains its meaning must be tied to other representations' meanings, such that we cannot necessarily invert the meanings of particular representations without downstream semantic effects. However, this result only demonstrates facts about intra-linguistic relations; we would never conclude that the strings 'This cat is white' and 'Are you an American citizen or a German citizen?' must be especially well suited to the aspects of the world that they represent. Surely these strings have their meanings only through historical accident. With sufficient creativity, we could unproblematically conceive of languages in which their meanings are

inverted fully such that they bear no special relations to their English meanings. Such languages would simply need to differ from English with respect to other symbols as well.[9]

By allowing for many "compensating" representational changes, we might similarly imagine more possible ways in which the spatial qualities of characters can vary their "semantic values". Thompson argues that some creature's squarish′ experiences of 2:1 rectangles could be veridical in the right circumstances, whereby some perceiver's "experiences of distance need not represent the very same qualities as [another perceiver's] experiences of distance" (2010, 180).[10] In response to the suggestion that the relation between tactile perception and visual perception provides reasoning for ascribing misperception to this seer, Thompson suggests compensating changes to this perceiver's tactile perception (2010, 178). Thompson, therefore, allows for some divergence among the spatial properties of veridical perceptions of objects caused by differences in visual apparatus. However, Thompson stops far short of FP by requiring that veridicality-preserving distortions preserve "spatial isomorphism", characterized as "an isomorphism between relations within spatial experience and relations among external spatial properties" (2010, 176). Yet, given that all of our sources of information about the external spatial properties of any particular objects are mediated by perception, in practice, this

[9] One might think facts require linguistic representations of suitable complexity, such that a complex string can't be translated to "P" without semantic loss. Yet, while semantically atomic in English, each "P" has a structure composed of infinitely many points. It's possible to map each semantically relevant component of any string onto a distinct component of "P", such that the strings offer equivalently complex semantic decompositions.

[10] Thompson adopts this example from (Hurley 1998). Chalmers endorses Thompson's judgment regarding the veridicality of this perceiver's perceptions, and suggests, correctly I think, that this conclusion leads to a general rejection of the possibility of lifelong spatial illusion (Clark and Chalmers 1998; D. Chalmers 2022).

requirement apparently amounts to preserving isomorphism with standard human shape perception, which is presumed isomorphic with "external spatial properties".

Thompson demonstrates that a perceiver with what we would consider a visual-spatial distortion may have compensating "distortions" impacting their tactile perception to preserve the sensible relations between their visual and tactile perceptions. To imagine more radical cases of character divergence, rather than undergoing the arduous task of considering how each aspect of some imagined experience might differ from the corresponding aspect of human experience, we might instead begin with a human wearing virtual reality goggles controlled by an attached camera such that the character of their visual experience is a direct function of the camera's input.[11] Our subject's visual character can be determined by some bijection from the character they would have experienced without the goggles. For any visual input i that causes a typical human to experience a visual field containing n squares, we could design goggles such that the goggled human inputting i experiences a visual field featuring n circles.

Now imagine creatures born with visual systems functionally equivalent to the augmented visual systems of humans in these goggles, such that when a normal human would somewhere in their visual field upon encountering input i, at least one of these creatures has n circles somewhere in their visual field when encountering i. By not stipulating that each square and circle must be in the same parts of their respective visual fields, we avoid the standard feasibility concerns, such as whether the parts of a

[11] Given my motivating question, I am focused here on VR instances in which character is directly correlated with the properties of the external physical world, as opposed to perception within virtual worlds as discussed in Chalmers 2016; 2022.

checkerboard would overlap. For all we have stipulated, a creature might, when confronted with a checkerboard, experience a visual field with all overlapping circles, no overlapping circles, some overlapping circles, circles that form a giant circle of their own, concentric circles, etc. So long as there are 64 identifiable circles somewhere in the visual field, the stipulated conditions have been met.

You might worry that basic geometrical worries already arise. For instance, let's say that your entire visual field is taken up by a large checkerboard. Then, Ally the alien takes your place and veridically sees everything that you saw, except that Ally's visual field has 64 non-overlapping circles, forming a large circle of their own, and no squares. We might ask: Does each circle represent each square of the checkerboard? If so, what does the funnily-shaped area in the middle of the 64 circles represent? Avoiding "representation" talk, we could alternatively instead say: Does Ally see each box of the checkerboard in a circle-ish′ way? If so, what, if anything, does she see in the manner of the funny shape between the circles? If the circles in Ally's visual field correspond to the squares in your experience, then — since you *only* saw checkerboard squares — Ally must see *more* than you, and thus cannot be seeing all and only the same objects that you saw.

This objection could arise from the following tempting principle:

> *Weak Isomorphism:* If S1 and S2 veridically visually perceive all of the same (parts of the same) objects, then, for each region R1 in S1's visual field, there is exactly one region R2 in S2's visual field such that R1 and R2 correspond to the same bit of the world.

However, the Normals and the Weirds already present a counterexample to Weak Isomorphism. A Normal and a Weird may both see only the triangular block, and yet, there is no particular point in the Weird's monochromatic visual field corresponding to any given point of the block.[12]

Even if Weak Isomorphism doesn't hold universally, you may think that complex perceivers must conform to Weak Isomorphism as a matter of feasibility. You might deny that creatures could have naturally evolved to see a checkerboard via a visual field containing 64 overlapping circles, no less circles forming a giant circle of their own. If such creatures' tactile systems were just like ours, the lack of appropriate correspondence between their visual and tactile perceptions may cause a functional disadvantage. However, just as Thompson describes divergent tactile experiences appropriate to divergent visual experiences, we can similarly imagine more radically divergent tactile experiences to correspond appropriately to radically divergent visual experiences. Along with VR goggles, imagine a permanent "skin suit" that detects objects in its environment via lasers and invokes tactile sensations in response. This suit also restricts the subject's motion so that, given the right sort of laser input, its inhabitants will have sensations such as those you have as you push up against walls or tap your desk. Consider a creature born functionally equivalent to a human wearing both the VR goggles and the bodysuit from birth. Since the camera and bodysuit are taking input from the same external objects, we

[12] You might think introducing the secondary perceiver is not relevant. All that matters is whether a part of the visual field corresponds to a part of the world. We could replace Weak Isomorphism with this parallel "Parts" principle. Yet, in attempting to determine whether or not some region of some visual field corresponds to some part of the world, we will need to perceive (or at least model) that part of the world in some way or another ourselves, and then use that experience of the world to evaluate the relevant perceiver.

can ensure they are well correlated so that the visual experiences "fit" with the tactile experiences. That is, if this creature's visual experience when looking at a checkerboard is like a normal human's visual experience of 64 circles arranged in a giant circle, then this creature's tactile experience of a checkerboard might be like a normal human's tactile experience of 64 circular things arranged in a circle. For achievement of the appropriate proprioceptive and self-awareness sensations, we could complement (or replace) the goggles and suit with instruments that directly impact the brain. If character supervenes on brain states, we could in principle make any action feel any way to our remodeled human. Then, we could again imagine a creature born that way. What it feels like for this creature to neatly cut up the checkerboard into its 64 individual squares and put them in a pile might be the same as what it feels like for a human to trace the outline of each of the 64 circles and then pile those up.

As I've been addressing the commonly discussed "spatial inversion" possibilities, I have focused on the case of a 64-square checkerboard seen via a visual field containing 64 circles. Yet, according to FP, we might have just as well considered a case in which the visual field contained 62 or 17 circles. Potential feasibility concerns for these cases might involve a commitment to the principle that two things that look similarly in one veridical perception must look similarly in all veridical perceptions. According to this "Similarity Constraint", you might think that all 64 boxes of the checkerboard must be seen in a similar way. However, that constraint fails for ordinary cases of perception. Take a triangular prism, a cube, and a square paper. Any two of these appear more similar than the third in some instances of veridical perception, even with the same parts visible. There is no reason to suppose that things that look similar to us should look similar to all other perceivers.

Notably, all objects have endlessly many properties, some more perceiver-centric than others, most of which we don't perceive. There are endlessly many similarities available to be perceived by different creatures with different interests and perspectives.

Returning to the question of mapping our experience of the checkerboard onto some circle-ish' perception of the checkerboard: there need not be a straightforward answer to which parts of our experience correspond to which parts of that experience.[13] We can answer questions about how changes from our perspective will correlate to changes in from others' perspectives, thus identifying how they track parts of the world that we track. Yet, some of the parts of the world that are important to us, like checkerboard squares, might seem like ad hoc mereologies to other perceivers, and vice versa with respect to parts of the world that they care about. Still, this does not mean that these creatures cannot successfully navigate the checkerboard as well as us. Since their odd-to-us visual experiences of the checkerboard will be matched with correspondingly odd-to-us tactile experiences and proprioceptive experiences, their navigation of the checkerboard need not involve ad-hoc strangeness identifiable within their perspective on the world. The strangeness only emerges when we attempt to "translate" their experiences into ours or vice versa. You might suspect that feasibility concerns due to structural differences will arise for this "translation", especially as we consider cases in which the complexity of a given perceptual experience seems misaligned with the complexity of the perceived objects. However, this "translation" can be as ad hoc as we wish. Our strangely perceiving creature is not impeded by the fact that how things look or feel to them is

[13] This could be put in terms of correspondence "to the world". The same concerns apply, as we still need to identify the bits of the world as we can perceive them.

radically different from how these same things look or feel to us. So long as their experiences fit together as well as ours and track as much information about the world, there is no reason to think that we could navigate the world more effectively than them.

Take the action of piling cubes to form a wall. This action consists of moving one's body in particular ways (the details, of course, depends on the sort of body one has). In principle, just as any visual or tactile input could produce any phenomenal experiences, undergoing such actions could produce any subjective sensations. You might think that building a wall of cubes would be difficult for a creature who has circle-ish experiences when looking at the cubes. However, if the act of piling up the bricks feels subjectively like, say, moving spheres into a row, such that how navigation proprioceptively feels is symbiotic with how the world looks and tactilely feels, such creatures could be as effective wall-builders as us. To make this work, the "translation" from their experiences to our experiences may seem rather strange and unnatural. For instance, giving up on the Similarity Constraint discussed above, two things that look very similar to us might look very differently to other creatures. A single, salient-to-us object may be an unremarkable jumble of parts to other creatures. This fact about the comparison between us and them does not impede their activities. We may suppose plausible restrictions on the similarity relations *within* the perspective of one creature. We might expect some general pattern of similarities such that, for instance, if the visual phenomenology of putting one brick on the wall is extremely similar to the visual phenomenology of putting another brick on the wall, then there will also be a similarity between the proprioceptive phenomenology of these actions. This principle is rather difficult to precisify given that, even for humans, similarity relations among visual experiences are far from a perfect guide to similarity relations

among proprioceptive or tactile experiences. However, so long as we have the freedom to make the bijection between imaged alien experiences and our own experiences as ad-hoc and unnatural-seeming as we like, we can ensure that similarity relations within the alien's experiences are at least as robust as our own. Thus, as far as the argument for FP goes, the precise requirements of internal similarity are unimportant.

Similarly, in imagining our functional equivalents, you might wish to impose some constraints for perception of apparently continuous changes over time. Yet, there is no reason to suppose that such constraints would cause a problem for FP. All we need to characterize about a possible perceiver S is that, upon seeing one particular set of objects (as they are at a stipulated point in time), S's perception has a particular character. The rest of S's perceptions, including S's visual perceptions of the given objects as they are at different points in time, S's visual perceptions of all other sets of objects, and all of S's other perceptions (and proprioceptive awareness) can be however would be most adaptive. And, after all, every visual phenomenal character could be continuously changed in infinitely many ways.

At this point, we find ourselves with respect to divergent shape perceivers just as we are to color inverts and as the Normals are to the Weirds. We lack non-circular justification for believing that our perceptions are more veridical than their functional equivalents. If we were to point out that these creatures' alleged perceptions violate Thompson's isomorphism principle, their philosophical representative could produce an equivalent argument demonstrating that human perceptions violate their isomorphism principle. Every claim about the structure of external properties will take for granted that

some representation of those properties shares their structure. If two perceivers experience the same tactile characters but different visual characters, one might offer the other an argument that their own visual characters more fittingly match the external world via a demonstration that relies on shared tactile experiences. However, without any such neutral representation, the claim that our perceptions are more veridical than those of functionally equivalent apparent perceivers has no grounds beyond blind faith in our superiority.

One might argue that the sort of tactile perception described above would be less direct than our own given that we tactilely perceive objects by bumping our body parts directly up against them, while these creatures' tactile perceptions are laser-mediated. However, such an argument implicitly depends upon the assumption that our visual perceptions are more veridical than these creatures' alleged visual perceptions. Our only justification for the belief that we are actually up against items while these creatures are not comes from our visual and tactile experiences; however, these creatures will similarly visually represent their own tactile perceptions to be more direct than ours. Just as Papineau concludes that "evolution has no doubt selected" the characters of our spatial experiences for representing corresponding worldly spatial structures, these creatures will conclude that evolution has effectively selected the structures of their experiences to represent worldly structures (2021, 111).

While I argue that our imagined creatures can perceive all objects we can perceive, they will likely talk about and care about different bits of the world than we care about. Consider creatures who, for instance, would, when visually confronted with all you are

seeing now, experience a character that you would describe as seeming like an abstract dot pattern. Such creatures would likely not bother to coin a term for your laptop, even if they encounter laptops frequently, given laptops' arbitrary role in their perspective on the world. Just as we see many entities, such as random mereologies of various bits of different objects, for which we lack special names or concepts, our ordinary objects might be perceived by creatures that do bother to name them. One might claim that there are self-evident restrictions on the character of veridical experiences of objects like laptops and pens. Siegel offers the following constraint on object perception: "If S sees o, then S's visual phenomenology differentiates o from its immediate surroundings" (2006, 434). As stated, this principle seems to fall prey to simple counterexamples, such as the case of an individual who stands extremely close to a huge red wall, such that their entire visual field is red′, and thus does not visually differentiate the wall from its immediate surroundings despite, it seems, seeing the wall. Interpreted to limit what it takes to specifically perceive o *as an object*, this principle is consistent with FP. Siegel's further suitability claim — "[I]f one saw an otherwise uniform expanse that was half blue and half green, that would not be phenomenology suitable for seeing Franco" (2006, 435) — might too be compatible with FP if we interpret "seeing Franco" to mean "seeing Franco as Franco".

FP is about character constraints on veridical perception of objects, but what about character constraints on veridical perception of properties? FP does not automatically rule out such constraints, given that veridical perception of objects does not require perception of all of their properties. For instance, one could veridically see a red box without perceiving the box as red. That being said, my argument for FP commits to the claim that, for any perceivable property p, one might perceive a property that necessarily coextends

with p via any character. Therefore, so long as properties are individuated intensionally such that distinct properties cannot necessarily coextend, my view is inconsistent with character constraints on property perception. If properties are individuated more fine-grainedly, my view could be compatible with such constraints. For instance, if one says that redness necessarily coextends with — but is distinct from — property p, then, without rejecting my argument for FP, they may claim that red′ experiences are required for perceiving o as red. Alternatively, without abandoning the intentional individuation of properties, one might preserve an intuition that seeing x "as red" requires redness′ by interpreting "seeing x as red" opaquely. On such a view, the sense in which some creatures don't see x as red, despite seeing x as p, even though redness is identical to p is like the sense in which one might not see that Superman has arrived despite seeing that Clark Kent has arrived even though Clark Kent is identical to Superman. Or, of course, one might simply reject intuitions like "seeing x "as red" requires redness" as mistaken, along with intuitions about character constraints on object perception. Any of these options is consistent with the view defended here.

While I have argued against an abstract Similarity Constraint, Similarity-Based concerns about concrete cases may linger. Imagine that you veridically see two qualitatively identical black cubes, side-by-side, each with exactly one side visible to you, floating in white space. Your visual field is symmetric along the veridical axis, with a black square on each side. Now, Alien Al takes your place, and looks at the same scene, perceiving everything that you perceive via a visual field just like yours, except that the right square is replaced with a black circle. Even if you accept that square surfaces can be veridically perceived circle-ishly, you might think that this case presents a further problem: If cube C2

looks circle-ishly′ from Al's perspective and cubes C1 and C2 are qualitative duplicates, then surely C1 should also look circle-ishly′ from Al's perspective! Isn't Al perceiving a difference where none exists?

First of all, recall that FP only guarantees that Al veridically perceives both cubes via the specified visual field, not that he specifically perceives one cube via the square and one cube via the circle. If we reject *Weak Isomorphism*, as I've advocated, we must be content to accept the possibility that there is no fact of the matter about how the sub-regions of your visual field correspond to the sub-regions of Al's visual field, even as we stipulate that you and Al perceive exactly the same objects. This possibility alone should suffice to quell worries about this case.

Still, setting aside the possibility of mere holistic translation, there are a variety of ways one might want to spell out a correspondence between Al's field and yours, depending on the omitted details about Al's perceptual faculties. For instance, perhaps the top halves of both shapes in Al's visual field "correspond" to the left square in your visual field, and the bottom halves of each shape in Al's visual field "correspond" to the right square in your visual field. Such a mapping preserves an intuitively appropriate similarity between how Al perceives the two cubes, thus evading the motivating worry. Yet, further cases could be cooked up where this sort of similarity-preserving mapping is not available.

Yet, in rejecting the Similarity Constraint above, I have argued that such similarity need not be preserved. So let us now consider a case in which Al *does* see one cube via the square and one cube via the circle. Must Al falsidically perceive a non-existent difference? Not necessarily. All differences in how the cubes appear may be still attributable to

differences between Al's respective perspectives on the cubes. Since the cubes are qualitatively identical, if we invert the positions of the cubes, Al's post-inversion visual field must be the same as his pre-inversion visual field. There are many ways to fill in the details of Al's vision to accommodate this. For instance, perhaps, square objects present more roundishly' when they are further right in Al's visual field, somewhat like how circular objects present to us more oval-ishly' when seen at certain angles. In this case, if we remove C2, and shift C1 gradually rightward, the left square in Al's visual field will gradually morph into a rounded shape, until eventually — when C1 reaches the original position of C2 — there is a circle on the right of Al's visual field exactly like the circle corresponding to C2 in the original setup. Given his lifetime of accommodation to this visual system, Al thus sees the cube's sameness in shape, despite their differing presentations, just as we recognize the sameness in shape of two circular tabletops presented at different angles.

I lack space here to develop a complete account of Al's perceptual systems. Still, to appreciate the plausibility of this case, we might imagine that Al, unlike humans and most of our evolutionary kin, is radically asymmetric — at least as we perceive him. Maybe Al's body is such that, as object o moves rightward in his visual field, which body parts can interact with o change such that whenever he tactically engages with the cube, its feel appropriately relates to its appearance. Perhaps different sorts of objects are easier to recognize and use when presented at different points on this spectrum, so accessing this range of perspectives benefited Al's ancestors. Still, even if it is useful for Al's shape perception to vary along the left-right dimension in this way, you may insist that Al perceives a difference where none exists. While the cubes differ in location, this difference

is already registered in Al's vision in the usual way. Al's perception thus seems to encode a difference in the cubes over and above their difference in location.

In response to this worry, consider another way the cubes differ: C1 is closer to Al's "circle-ish side" — the side on which square surfaces appear (and feel) circle-ishly′ — while C2 is closer to Al's "squarish side". This difference between C1 and C2 only contingently correlates with their difference in location with respect to Al, given the possibility of Al's anatomical inversion. Of course, we could identify a similar difference between your perspectives on the cubes, focusing on the different relations C1 and C2 respectively bear to your left and right eyes. However, the near-perfect symmetry of your visual system explains why this difference is effectively irrelevant to your experience. In contrast, due to the dramatic difference between the sides of Al's perceptual system, the side of his visual system from which Al perceives an object impacts character beyond determining which side of his visual field present that object. Al simply occupies a greater diversity of perspectives than we do.

We can thus understand Al's case as a slight variation on the cases of Character Divergence introduced in Section 2. The standard cases motivating this chapter take the following form: S1 and S2 see objects W via radically different characters. If both perceptions are veridical, and the seen objects are identical, then what could explain this character divergence? In short, my answer has been: facts about the difference between S1's and S2's respective perspectives on W. I have argued that perspectival facts necessarily determine character in conjunction with facts about the nature of perceived objects. Different perspectival facts can account for different characters, even when the

perceived objects are held constant. In the case of Al, instead of comparing two subjects' perspectives on the same object, we compare one perceiver's perspectives on two objects. Mirroring the above form, Al's case can be characterized as follows: Al sees C1 and C2 via radically different characters. If both perceptions are veridical, and the seen objects are (qualitatively) identical, then what explains this divergence in character? I answer: facts about the difference between Al's perspective on C1 and Al's perspective on C2.

Finally, one might worry that my arguments about the subjectivity of similarity relations across space will overgeneralize to include similarity relations across time, leading to a more radically revisionary view. For instance, let's say S watches a static scene for one minute, receiving constant visual input across this period. However, 33 seconds in, S's experience undergoes a dramatic shift in character. From our perspective, it may seem as if this change is caused by a change in S's visual system. That is, after 33 seconds, aspects of S's visual system alter such that the same visual input now leads to a different brain state, causing S to have a different experience despite apparently watching a static scene. However, if we invert the scenario such that S is watching us watch the scene, it may have seemed to S as if our visual system had abruptly changed 33 seconds in, such that, although we started receiving different input, our brain state (with respect to our visual experiences) remained the same. You may think that my arguments lead to the claim that both S and we may see this scene veridically. If this is the case, we get a vastly more radical view, leading to a version of FP that applies to lifetime experiences and trivializes perceptual reliability, and eliminates the possibility for any substantive notion of veridical perception.

However, we have good reason to reject this more radical version of FP. As the case is described, according to S's own perspective, S's visual system, along with the rest of S, will have also changed during this period. Perhaps, we could construct a version of the case in which each creature of this species undergoes a certain change in perspective at a set point in time, such that S anticipates and understands the change, and, perhaps perceives veridically prior to and after the change. Or, maybe S is ignorant and doesn't realize that he has changed. Still, it is within S's power — or, at least, within the power of someone with S's capabilities — to scientifically discover this change within the S perceptual systems, and to discover that it is in fact him and not the world that has changed. No symmetric evidence is available to us suggesting that we've changed.

I have not proven that there could not be some abstract, true principle entailing constraints on how a given set of objects could be veridically perceived by any well-functioning creature, regardless of the rest of the subject's perceptual systems. I have made the case that we cannot presume there is such a principle, that we have reason to think there is no such principle, and that simple candidates for such a principle will fail. If such a principle exists, its identification would be of great value to philosophy. Either way, the accepted inference from the difficulties involved in systematically "inverting" human visual spatial perception to character constraints on veridical perception must be abandoned. Furthermore, without non-circular justification for perceiver-independent character constraints, we should continue to treat more radical character divergence as we treat everyday character divergence.

6. Conclusion

I have argued that any visual character is capable of being the character of a veridical perception of any visually perceptible set of objects. Given that everyone already accepts some character divergence among veridical perceptions of the same objects, accepting FP only amounts to changing the degree of divergence permitted, as opposed to an in-kind shift. Accepting FP should not threaten our confidence in the veridicality of the commonsensical or standard scientific truths we learn about the world via perception. Still, with respect to philosophical and metaphysical investigation, FP has radical consequences for attempts to ascertain universal truths about the world's structure. If FP is correct, then the character of each of my experiences (independently of it being specifically *my* experience) is compatible with every possible subset of properties independent of me. Furthermore, this conclusion applies equally to my perceptions of what perceivers like me are like and how sensory apparatuses like mine are constructed. Thus, when I perceive that something seems a certain phenomenal way w to a perceiver like me, I perceive only that some bit of the world seems way w to *some* perceiver, namely *me*. And, if my account is correct, *any* visually perceptible set of objects is such that it would seem way w to some possible perceiver.

References

Byrne, Alex. 2001. “Intentionalism Defended.” *The Philosophical Review*: 199–240.

Campbell, John. 2002. *Reference and Consciousness*. Clarendon Press.

Chalmers, David. 2004. “The Representational Character of Experience.” In *The Future for Philosophy*. Clarendon Press.

—— 2016. “The Matrix as Metaphysics.” In *Science Fiction and Philosophy*, 35–54. Hoboken, NJ: John Wiley & Sons, Inc.

—— 2022. *Reality+: Virtual Worlds and the Problems of Philosophy*. Penguin Books.

Clark, Andy, and David Chalmers. 1998. “The Extended Mind.” *Analysis*: 7–19.

Dretske, Fred. 1969. *Seeing and Knowing*. Routledge & Kegan Paul.

Hurley, Susan L. 1998. *Consciousness in Action*. Harvard University Press.

Levine, Joseph. 2018. *Quality and Content: Essays on Consciousness, Representation, and Modality*. Oxford University Press.

Logue, Heather. 2012. “IX-Why Naïve Realism?” *Proceedings of the Aristotelian Society*: 211–37.

Mendelovici, Angela. 2013. “Reliable Misrepresentation and Tracking Theories of Mental Representation.” *Philosophical Studies*: 421–43.

Molyneux, W. n.d. “1688, Letter to John Locke, 7 July.” *The Correspondence of John Locke*.

Montague, Michelle. 2013. “The Access Problem.” *Phenomenal Intentionality*, 27–48.

Papineau, David. 2021. *The Metaphysics of Sensory Experience*. Oxford University Press.

Pautz, Adam. 2021. *Perception*. Routledge.

—— 2023. “Naïve Realism v Representationalism: An Argument from Science.” *Contemporary Debates in Philosophy of Mind.* Hoboken: Blackwell.

Peacocke, Christopher. 1983. *Sense and Content: Experience, Thought, and Their Relations*. Clarendon Press.

Siegel, Susanna. 2006. "How Does Visual Phenomenology Constrain Object-Seeing?" *Australasian Journal of Philosophy*: 429–41.

Thompson, Brad. 2010. "The Spatial Content of Experience." *Philosophy and Phenomenological Research*: 146–84.

Chapter 2: The Noumenality Myth

ABSTRACT: There is a long-standing debate about whether our perceptions transparently reveal the nature of worldly objects. Another debate is about whether the objects of perception place any constraints on the phenomenal character of veridical perception. There is a consensus that these debates are related: endorsing fully transparent perception requires character constraints on veridical perception. This chapter attacks that consensus. The consensus requires positing a family of properties whose metaphysical status is much more peculiar than is generally appreciated. These "noumenal" properties are allegedly essential to explaining the veridicality of our perceptions, yet no clear explanation is available for how we can learn about them or why we should postulate them. I argue that they do not exist.

1. Introduction

You see a red cubical box, well-lit on the table before you. Your vision is working perfectly, and you see the cube as it is. How would you describe what your experience is like? (In alternative jargon, what is the "phenomenal character" of your experience?) According to many philosophers, you can't do better than describe what the *cube* is like. Searching for any "intrinsic qualities of experience" comes up empty: all you find are intrinsic qualities of the cube, and how it is related to other items in the scene, and to you (Harman 1990). Here is how Campbell puts the point:

> [T]he 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. (2002, 116)

Seeing the red cube is not like reading about the cube. Reading the sentence "There is a red cubical box on the table before you" makes you aware that there is such a cube. But in describing what your experience of *reading* about the cube is like, you would mention

features of the words — that they are black and in Cambria 11 pt., say. According to Campbell, nothing analogous holds for your experience of seeing the cube. Stated more generally, Campbell's view is this:

> *Transparency:* The phenomenal character of veridical perception is constituted by (aspects of) the actual layout of the perceived scene.[14]

Let's now turn to another question. Imagine that Martian Martia, whose perceptual systems are functioning as Martian evolution has designed them to function, also perceives the red cube on the table. However, Martia's experience is phenomenally exactly like your experience would be if you were to see a green sphere on the table. As we can put it: Martia's experience of the red cube is greenish and spherish, whereas your experience of the red cube is reddish and cubish. Could Martia's experience of the red cube be veridical? Well, why not? The case of Martia seems possible, and it seems unacceptably arbitrary to pick your experience of the cube as the only veridical one. Reflections on examples like this one lead to the following thesis, defended in Chapter 1:

> *Arbitrariness:* For any phenomenal character, and any scene, there is a possible veridical perception of that scene with that phenomenal character.

(Arbitrariness could be qualified in various ways — for our purposes, this won't matter.)

[14] Campbell's talk of 'constitution' can be confusing: A simple illustration of the view is this. Concentrating exclusively on color, the phenomenal character of a veridical experience of a red object = the property of perceptually acquainting the subject with redness.

Transparency and Arbitrariness are usually held to be incompatible: we can have one, but not both. And the argument for this incompatibility can seem decisive. Here are you and Martia, looking at the red cube. If Arbitrariness is true, we may suppose that your experience and Martia's experience are both veridical. By Transparency, the phenomenal character of your experience is constituted by aspects of the scene — including, presumably, the red color and cubical shape of the box. By the same token, the phenomenal character of Martia's experience is constituted by the green color and spherical shape of the box. But the box can't be both red and green, or cubical and spherical! Contradiction: either Arbitrariness or Transparency must be rejected.

In what follows, I argue that Arbitrariness and Transparency are not in conflict after all. The argument that they are tacitly relies on postulating superempirical "noumenal" properties that set the standard of veridicality, a postulate that we have no reason to make. Once we reject this postulate, the way is clear to endorse both Arbitrariness and Transparency. Section 2 introduces the tacit assumptions in the debate. Section 3 offers an argument against the noumenal postulate. Section 4 considers potential scientific justifications for the noumenal postulate. Section 5 addresses a key objection. Section 6 considers how the metaphors of perception perpetuate the noumenal postulate. Section 7 contends with the upshots of abandoning noumenality.

2. Transparency, Arbitrariness, and Incompatibility

Transparency is primarily defended on the basis of phenomenal reflection; it is often pointed out that attempting to reflect on the character of your perceptual experience of,

say, a red cube, seems to amount simply to trying to reflect on what the red cube is like.[15] Another motivation for Transparency derives from an attempt to avoid the skeptical challenge facing some forms of Indirect Realism — i.e., if all we directly access are sense-data, how do we know that worldly objects exist? According to the view that perception is not transparent, the perceptual characters of our experiences track worldly properties but do not reveal the true nature of such properties. This kind of sense-data view was famously defended by Russell:

> It has appeared that, if we take any common object of the sort that is supposed to be known by the sense, what the sense *immediately* tells us is not the truth about the object as it is apart from us, but only the truth about certain sense-data which, so far as we can see, depend upon the relations between us and the object. Thus what we directly see and feel is merely 'appearance', which we believe to be a sign of some 'reality' behind. (1912, 16)

Not many contemporary philosophers endorse anti-Transparency accounts of perception, with Papineau offering a notable exception:

> It is natural to think of sensory experience as in some sense providing a bridge between the mind and the world beyond. Both naïve realism and representationalism support this natural thought in their different ways, by building worldly facts or properties into the fabric of experience itself. The qualitative view [Papineau's] rejects any such worldly involvement in experience. It says that conscious sensory properties are sufficient unto themselves, enclosed within an internal subjective realm, and have no essential connection to anything beyond. (2021, 7)

Outside philosophy, anti-Transparency is perhaps more popular. A recent book by cognitive scientist Donald Hoffman offers an even more radical rejection of Transparency:

> [T]he probability is zero that we see reality as it is. This theorem applies not just to taste, odor, and color, but also to shape, position, mass, and velocity — even to space

[15] For this sort of claim, see: Brewer (2007), Byrne (2001), Campbell (2002), Logue (2012), Harman (1990), Tye (2003).

> and time. We see none of reality as it is. The reality that prompts you to create an experience of a tomato, the reality that exists whether or not you see a tomato, is nothing like what you see and taste. (2019, 60)

While Transparency characterizes the relationship between phenomenal character and the qualities of the perceived world, Arbitrariness concerns the relationship between character and perceived entities. When we ask whether your perception of the box is *transparent*, we are asking whether the phenomenal character of your experience of the box is constituted by qualities of the box or merely signifiers of qualities of the box. When we ask whether your perception of the box is *arbitrary*, we are asking about the range of potential characters of veridical perceptions of the box. Once Transparency is accepted, Arbitrariness is generally presumed false, often without further argument. The alternative reasoning also occurs: if we could have had different perceptual systems that represented the world just as well via radically different characters, then none of these characters could be transparently revelatory. I argue that these moves are unjustified.

The alleged tension between Transparency and Arbitrariness is often characterized as an issue of dependence. As Pautz asks, in introducing the problem of perception, "How can perception be both internally dependent and externally directed?" (2021, 1). How can perception be both based on arbitrary facts about perceivers and transparently reflective of the external world? The central disagreement between those who embrace perceptual transparency and those who embrace perceptual arbitrariness is framed in terms of whether and to what extent the phenomenal character of veridical perception depends on, or is determined by, perceived external objects as opposed to the subject.[16] However, this

[16] For example, see Pautz (2021), Logue (2012), or Hoffman (2019).

dependence issue is a red herring. Even if which properties the box has does not depend on you and even if the character of your perception were entirely determined by perceived properties of the box, which of the box's properties you perceive may depend on you and your perspective. For instance, you might see the redness of the box but not its magnetic field, while a shark might see the latter but not the former. Facts about you determine that your perception is reddish rather than however magnetic fields seem to sharks; yet, this fact does not threaten the transparency of your perception. After all, the box has both redness and a certain magnetic field.

However, what about cases without an obvious category difference among apparently perceived properties? If you and I both perceive the same three sides of the same cube floating in space, the sizes and shapes of the regions of our visual fields may differ on account of our respective spatial positions. Here, we may presumably both veridically perceive the cubicality of the box. Still, we are each perceiving the cubes in relation to different points in space, and our phenomenal characters appropriately differ. Again, no conflict for Transparency arises because of the compatibility of our characters. Brewer, another fervent defender of Transparency, describes perceptions as "conscious acquaintance with particular mind-independent physical objects, from a given point of view, in a particular sense modality and in certain specific circumstances of perception" (2007, 118). Variations in points of view, sense modalities, and other circumstances allow for the cube to seem differently to different perceivers, without threatening Transparency. Your, my, and the shark's perceptions may reveal how the cube "really is", but we each only get part of that story.

The trouble arises when the stories allegedly conflict. Consider Martia, who perceives the box via an experience that is phenomenally like a typical human experience of a green sphere. We can call this a green-spherish experience. (I'll call an experience a "Pish" experience if it is phenomenally like a typical human experience of P. To Pishly perceive o is to perceive o via a Pish experience.) If Martia were human, you can safely conclude that she (or you) is misperceiving. A thorough examination of her (or you) should eventually reveal the cause of this perceptual failure. Yet, what if Martia is a Martian, with a visual system completely unrelated to your own? Knowing nothing about this system, can you conclude that if your perception is veridical and transparent, hers is not? It is generally presumed that the respective characters of your and Martia's perceptions are incompatible, such that both cannot be transparently revealing aspects of the exact same parts of the world. Crucially, this incompatibility claim is not about how the box might seem to different creatures, but only about how the box might seem to creatures who perceive the box *veridically.* There *could* be an alien perceptual system like Martia's, insofar as there could be a creature with eyes that receive light reflected off of the same bits of the world as you see, who then, per some reliable processes, has a green-spherish experience.[17] That much is fairly uncontroversial. The controversial claim is that such an experience could reveal the nature of the box. This takes us to a third thesis, which is required for the incompatibility between Arbitrariness and Transparency:

[17] You might dispute that this perceptual system could facilitate efficient navigation. This will be addressed in Section 4.

Incompatibility: Some pairs of phenomenal characters are constituted by incompatible features of perceptible scenes.

Without Incompatibility, Arbitrariness and Transparency are perfectly compatible. No matter how many arbitrarily different perceptions there are of a given set of objects, all of those perceptions can be transparent so long as the relevant objects have all of the respective properties corresponding to all of the perceptions. This is how I will argue things are. The problem only arises if we say that some of those perceptions, due to their phenomenal characters, must be attributing incompatible properties to the objects. I will argue that Incompatibility is false: objects do have a sufficient variety of properties to be revealed by arbitrarily different perceptions. I will argue that the connection between perceptions and the properties of objects is tighter than generally assumed.

Incompatibility is not about what empirical evidence anyone might receive. Instead, Incompatibility tells us how to assess that evidence. It is not about who will perceive what, but which perceptions are the correct ones. Incompatibility reasoning tells us that Martia's experience might be as described, but this perception could not veridically and transparently reveal the box, so long as our perception is veridical and transparent. For us to know that certain character sets are incompatible, the properties that we learn about via perception must put us in a position not only to predict and evaluate the experiences of particular perceivers, but also to evaluate the perceptions of all possible perceivers — e.g. to determine that no possible perceiver's cubish perceptions of a sphere could be transparent and veridical. In the next section, I argue that such properties are mythic.

3. The Noumenal

You are looking at what seems to be a shiny red cube, which you name 'Cubey'. In a skeptical frame of mind, you ask, *Is Cubey really as it seems? Or, is Cubey really a shiny, red cube?* Perhaps Cubey is only a hologram. Perhaps Cubey is not remotely cubical and/or red, and this is some optical illusion of angles and lights. Perhaps Cubey is oblong with invisible sides. You examine Cubey closely from many perspectives to see whether it might be some trick object. You do some stacking, throwing, and sliding, and Cubey seems to behave like a standard cube. You enclose Cubey into a tightly fitting cube-shaped box to ensure Cubey doesn't have any invisible parts. You use Cubey to construct complex machines that require a cube to function. At the lab, you examine Cubey under a microscope to learn about its crystalline structure. Still, you wonder whether the testing devices are faulty. You test all devices, and all seems well. Still, you can worry about these tests, or, more generally, whether you are in some Cartesian-style skeptical scenario. As it happens, you have not been plunged into a VR machine, you are not sleeping, and your perception is as reliable as anyone's, including those who coined 'cube'. Even so, there are facts about Cubey's properties of which you are ignorant, given your limited perceptual capacities. Some of these facts about Cubey are directly tied to how Cubey seems from perspectives that you can't access, such as how Cubey might seem to a radically different sort of perceiver.

Following Shoemaker, we can classify "phenomenal properties" — sometimes called "appearance properties" — as dispositions to appear in certain ways to perceivers with certain kinds of perceptual systems (Shoemaker 1994, 2000; Egan 2006). How Cubey looks to someone with a visual system like yours located where you are with respect to Cubey,

under the current circumstances would be a phenomenal property of Cubey. Another phenomenal property is how Cubey feels to someone with a tactile system like yours when they touch Cubey's sides in a certain way. Others involve how Cubey seems to other sorts of perceivers, possible and actual, in a variety of circumstances, possible and actual. For instance, the fact that Cubey would spherishly appear to someone with a visual system like Martian Martia's is a phenomenal property of Cubey. None of these properties is a serious candidate for cubicality; any is obviously compatible with Cubey not being cubical. If Cubey looks paradigmatically cubical to humans from one angle, it might be hollow on the other side or have unseen parts. If Cubey feels paradigmatically cubical to humans in one position, it might be a tactile illusion or have unfelt parts, etc.

However, combining and iterating these first-order phenomenal properties of Cubey, along with phenomenal properties of other things related to Cubey — such as the other objects Cubey engages with, the lasers used to test it, the eyes used to see it — gives rise to ever more complex properties that become harder to distinguish from cubicality. Eventually, iterating these properties gives rise to infinitely complex properties at which we can only gesture. For instance, in the story above, you didn't just feel Cubey, you also used testing devices to examine Cubey and then examined the testing devices. If we first consider how Cubey looks through a microscope or fits into a cubical box or feels to you, we can also learn more about Cubey by learning more about the microscope, the box, and your perceptual systems. All of these entities have their own infinite sets of first-order phenomenal properties. Combining them with Cubey's first-order phenomenal properties gives rise to more interesting properties of Cubey. For instance, we can consider the way that Cubey feels in various circumstances to a human whose brain looks a certain way

through an MRI machine that itself seems to behave certain ways when tested, etc. Iteration *ad infinitum* gives rise to infinitely complex properties that go far beyond what we could, even in principle, even with infinite time, translate into first-order phenomenal properties.

All such properties together constitute what we can call Cubey's *phenomenal profile*. To understand the scope of this profile, consider some of the questions that could be answered upon learning about parts of it: How would Cubey look to an alien perceiver with XYZ visual system? How would the results of an XYZ visual system's ABC lab test look to me? How would X part of the XYZ visual system feel when touched in way w by an alien perceiver with QRS tactile system? How would the results of a Y test for a brain of the TUV tactile system look to me? How would the test machines of the aforementioned tests look under a microscope? All of this and infinitely more is part of Cubey's phenomenal profile. No matter how much of this information you learn about Cubey, there will always be infinitely more information about Cubey's phenomenal profile left unknown. Many parts of this profile could help us negatively answer our initial skeptical question 'Is Cubey really how it seems?' For instance, perhaps, despite looking cubical from our initial perspective, Cubey is not a cube but instead the sort of trick object you might find in a museum of illusions. When we look at Cubey from other sides or grab Cubey, we realize that our initial perception was misleading and/or illusory. Perhaps, when we prod Cubey or look through a microscope, we will realize that Cubey is only a hologram. Or perhaps we could learn about abnormalities occurring in our own eyes or brains while we experienced Cubey or elaborate tricks causing our experiences to mislead us. Perhaps, had we been functioning normally, Cubey would have seemed completely different to us. Perhaps by examining the

lights in the room or by looking at the microstructure of Cubey, we will learn that Cubey isn't really red, and our first perceptions were affected by a trick of the light.

However, is Cubey's entire phenomenal profile (which could never be fully described) enough to determine that Cubey is definitely a cube? We have some idea of what it is to look or feel to humans like a cube. We also have a good idea of the sorts of evidence that would show you that Cubey is not actually a cube, despite looking or feeling like a cube in some instance. Without this understanding, we could never be justified in ever concluding that anything is more likely to be a cube than anything else. We can say that Cubey is *cubical** iff learning about Cubey's phenomenal profile would justify the belief that Cubey is cubical, and no further part of Cubey's phenomenal profile cancels out that justification. For instance, in the story above, Cubey is necessarily a cube* given that Cubey seems like a cube according to all of the infinitely many tests we could do, and, as we have stipulated, that there is nothing relevantly abnormal about our visual systems or other apparatus as we examine Cubey. Basically, Cubey is a cube* because Cubey's phenomenal profile conforms to our concept of what cubes seem like. (For now, we can stay neutral about how cubicality* relates to cubicality, although I will argue that they are the same.) Even if Cubey looked (to humans) deceptively like a sphere from a particular angle, such that an observer might be justified in falsely classifying Cubey as a sphere, Cubey would still be a cube*, not a sphere*, given that further empirical evidence could demonstrate the illusoriness of the spherish perception. Of course, Cubey's phenomenal profile *does* include Cubey's looking spherishly to Martians; yet, if this prevents us from concluding that Cubey is a cube, then nothing can be a cube. After all, for any object, there could be some visual system that takes in visual data from that system and gives rise to a spherish experience.

More generally, for any property P for which any actual or possible perceiver could, in principle, gain empirical evidence, we can identify a P* property such that o is P* iff o's phenomenal profile contains justification that o is P and no other part of o's phenomenal profile nullifies that justification. Essentially, P* is identical to P with respect to its phenomenal footprint. Any property that has any phenomenal footprint — and thus any property which one could study via the scientific method — can be given a starred counterpart that is entirely determined by an object's phenomenal profile, even if the original property isn't straightforwardly perceptible. So long as the claim that o is P entails something about how the world might seem from some possible perspective, occupied or not, accessible to us or not, there is a p* property that is empirically equivalent to p. Thus, the realm of phenomenal profiles ought not to be confused with Sellars' "manifest image", as contrasted with the "scientific image", given that both the "scientific image" and "manifest image" are relevant to the phenomenal profiles of objects (1962).

By definition, any empirical evidence that Cubey is cubical is evidence that Cubey is cubical* and vice versa. Any claim about cubicality that has any relevance to our perceptions or scientific investigations applies equally to cubicality*. So is cubicality* cubicality? Well, here is another, Incompatibility-based claim about cubicality:

> *Spherish Cube Incompatibility:* If Cubey is cubical, then Martians' spherish perceptions of Cubey are falsidical, or at least non-transparent.

You might think this claim flows straightforwardly from the concept of cubicality. However, if "cubical" is taken to mean "cubical*" here, this conditional is rather mysterious. The claim that Cubey is cubical* is a claim about how Cubey seems/would seem to all kinds of

perceivers. One part of being cubical* is seeming cubishly from standard human perspectives; another part of being cubical* is looking spherishly from standard Martian perspectives. This fact about how Cubey seems from different perspectives cannot alone justify the further claim that the way that Cubey seems in these circumstances is correct and the way that Cubey seems in those circumstances is wrong, unless the descriptions of the various perceptual circumstances themselves can justify the claim that some circumstances are more ideal than others — that is, unless we have some independent reason for thinking that humans are superior to Martians qua perceivers. (Potential explanations in this vein are considered in Section 4).

This point can be generalized. For any property P, it will not follow from the fact that some o is P* that o can never be transparently, veridically perceived via perceptions with certain phenomenal characters. That would be akin to saying: P things are perceived way 1 in circumstance 1, way 2 in circumstance 2, etc. *Therefore,* P things can never be transparently, veridically perceived in way 2. Properties that are fully determined by phenomenal profiles (henceforth "phenomenally determined properties") cannot get us Incompatibility. Incompatibility is not an empirical or predictive claim; it instead tells us only how to assess the empirical evidence.

Of course, cubicality* and sphericality* are phenomenally determined incompatible properties. Yet, there is no reason to suppose that Martia perceives Cubey as spherical*. Sphericality* is constituted by an infinitely large set of phenomenally determined properties, none of which include appearing spherishly from the perspective Martia occupies. We can describe S's perspective on o as all of the facts about S (about S's brain,

sensory apparatus, location in space, etc.) that affect the character of S's perception of o. Phenomenal spherishness is only specially linked to sphericality* relative to our perspective. It is a genuine aspect of sphericality* that spheres* look spherishly to humans in certain circumstances, and thus, it is plausible that humans normally see things that seem spherishly as spheres*. Yet, it would be absurdly arbitrary to suppose that Martia perceives Cubey as spherical*. Similarly, it is a genuine aspect of cubes*, albeit one we are not acquainted with, that they seem spherishly from normal Martian perspectives, and thus, Martia may also see Cubey as cubic*. If this is the case, there is no incompatibility, as you each simply perceive cubicality* from a distinct perspective with a different aspect of cubicality* featuring in your experience, just as two people may see the same cube despite seeing different sides. Given what has been said about Martia so far, it is also possible that she doesn't see Cubey's cubicality*, but instead sees a different phenomenally determined, non-shape property of Cubey, perhaps one that humans don't see. Either way, we don't get incompatibility issues. Thus, if it follows from the concept of cubicality that transparent spherish perceptions are incompatible with cubicality, then the attribution of cubicality must go beyond the attribution of cubicality*.

Incompatibility requires augmenting phenomenally determined properties, which determine how things seem, with further "noumenal" properties that determine the veridicality of those seemings. This may seem like an obvious step. Surely, there must be facts about how things *really are* in addition to facts about how things seem. If Cubey *really is* cubic, then your perception is veridical and Martia's is falsidical, or at least opaque — or so the thinking goes. Thus, truly being a cube in the sense required by Spherish Cube Incompatibility requires something beyond cubishness*. Yet, it is unclear what is required

for a property to be sufficient for characterizing what things are "really like". After, Cubey doesn't just *seem* cubical*; Cubey *is* cubical*.

Being cubical* and seeming cubical* can come apart. For example, perhaps your brain goes haywire and you have a cubish experience when looking at a sphere. This sphere might falsidically seem cubical* to you in this moment without being cubical*. Yet, the falsidicality of this perception can be explained by facts about you as a perceiver, combined with facts about Cubey's phenomenal profile. Some claims about how something *should* look/sound/feel/etc. are about the perspective that some particular perceiver *should* occupy. For instance, we can truly say that the sphere should not look cubishly to *you* because the sphere does not look cubishly from the perspective of a well-functioning human. The sphere should look spherishly to you because you should be occupying the perspective from which the sphere *does* look spherishly. Here, the normative sense in which your eyes and brain should behave in a certain way is the same as the sense in which your heart should pump blood. Yet, this claim relies on facts about your particular functioning. The negative assessment of how Cubey looks to you in some particular instance can stem from the normativity of how you should function. However, Incompatibility requires blanket negative assessments of any perception of Cubey with a certain character, regardless of the circumstances of perception. This assessment is generally presented as arising from the nature of cubicality rather than the nature of the perceiver, thus justifying the unlimited scope of the assessment. We can say that any perceiver, *qua* perceiver, should perceive things as they are; yet, this just takes us in a circle without helping to justify the claim that cubes* cannot be seen "as they are" spherishly. If Cubey really does have this property of only being veridically, transparently perceivable in

certain ways, this property is not phenomenally determined. This alleged noumenal cubicality must then be postulated in addition to cubicality* and all phenomenally determined properties. I claim that we have no evidence for this noumenal property. This position ought not to be confused with the position that things seem cubical but nothing is "really" cubical. Instead, we should accept that cubicality* is sufficient for cubicality; we ought to abandon the further requirement of noumenal cubicality. I think that being cubical* is being cubical, and that only metaphysical confusion, antiquated superempirical assumptions, and confused metaphors lead to the notion that cubicality requires something that goes beyond what is phenomenally determined.[18]

Anyone who accepts that things can seem different ways to different perceivers will have to accept that there are some phenomenally determined properties, even if they deny the importance or fundamentality of those properties. As we try to explain how things seem to us and theorize about how things seem when we aren't directly observing them, we postulate more phenomenally determined properties to fill in our worldview and create coherent, explanatory models. Yet, the further postulate of noumenal properties does not help us to predict or explain the empirical facts — as the phenomenally determined facts fully determine the empirical facts — but instead only provides assessments of those facts. I argue that we lack evidence for the noumenal and thus should eliminate the noumenality postulate. As there is nothing inherently contradictory about noumenal properties, we cannot prove their nonexistence. Yet, if we have no evidence of any kind for their existence

[18] I am concerned here with the cubicality property that is meant to apply to worldly, perceptible objects. Perhaps, in other contexts, including geometrical contexts, 'cubicality' refers to a property of representations themselves or to the property of being reliably represented cubishly or to some abstract mathematical entity.

and no reason to suppose they do exist, we can conclude they do not exist or at least decline to positively believe in them.

I have argued that cubicality *is* cubicality*, and we are not justified in positing noumenal cubicality in addition to cubicality*. One sort of Incompatibility defender agrees with this: You might agree that cubicality* and cubicality are the same property, such that the fact that Cubey is cubical* entails that Cubey is cubical. Then, you might defend Incompatibility by insisting that Spherish Cube Incompatibility follows from the concept of cubicality. However, as I've argued, the claim that spherish perceptions of cubes cannot be transparently veridical does not follow from the definition of 'cubicality*'. Still, accepting this conceptual difference between 'cubicality*' and 'cubicality', you might allege that these concepts simply contain different modes of presentation of the same property, like 'water' and 'H20'. However, the challenge of justifying the hypothesis that cubicality* is identical to a cubicality property that has the noumenal feature of providing character constraints on veridical perception is much the same as the challenge of justifying the hypothesis that cubical* things also have a different, noumenal property that provides those constraints. Either way, the difficulty resides in the inference from information about phenomenal profiles to information about character constraints on veridical perception; no empirical evidence can justify this leap and there does not seem to be any conceptual argument available that does not simply presuppose a normative superiority of humans to other perceivers. If it is coherent to suppose that you are veridically perceiving Cubey and Martia is not, then surely it is coherent to suppose the opposite is true. It is hard to imagine a logical obstacle that would apply to one case and not the other. Yet, if these are both genuine possibilities, then cubicality* cannot be identical to any cubicality property that

necessitates that spherish perceptions of cubes cannot be transparently veridical. After all, it is possible that Cubey, despite being a cube*, is veridically, transparently perceived by Martia's spherish perceptions.

4. The Scientific Perspective

For Descartes, plausibly the father of contemporary philosophy of perception, the mind of God grounds the noumenal. How things seem in the mind of God constitutes how they truly are, how they look *sans* any worldly window of perception (1681/1984). Since God is inherently perfect, God's perspective offers the ultimate standard against which all perspectives can be evaluated. Furthermore, Descartes only concluded that he could trust his phenomenal experiences as evidence of the noumenal because he was convinced of a logical proof that this perfect, and therefore non-deceptive, God would ensure that how the world seemed to humans aligned with how it *should* seem. While Descartes' question — do my perceptions reflect how the world is? — has survived through the ages, his metaphysical postulate of a perfect Godly mind that grounds the noumenal has been largely disconnected from the discussion. Given a secular, scientific worldview, it is mysterious how we could justify positing the noumenal and its alleged relationship to our perceptual evidence.

In referencing the noumenal, many turn to science. There is a common notion that science allows us to learn about how things are, independently of any way they might seem. The elusive, final scientific model replaces the mind of God as grounds of the noumenal. The following narrative is common: We begin with ordinary perceptions of the world.

Then, our scientists go out and investigate that perceived world, tell us about its true properties, and put us in a position to ascertain the extent to which our perceptions are veridical and transparent. However, the scientific method requires that even the most rigorous science ultimately relies on perceptions of the world; thus, the question remains how any given set of perceptions relate to a postulated noumenal realm. While the perceptions achieved through careful scientific investigation might be especially useful, informative, explanatory, etc., it is unclear what justification could be offered for the claim that they are especially likely to be connected to the noumenal—that is, that they can somehow teach us not only how things seem from various perspectives but also how they *should* seem to all perceivers.

Mendelovici says that "if we want to know whether color realism is true, presumably we should check the surfaces of objects for properties that could plausibly qualify as colors" (2013, 438). Of course, one way of "checking" the surface of an object is looking at it in good lighting. When we do this, it is trivial to identify "properties that could plausibly qualify as colors", given that our color concepts were coined in such circumstances. Yet, these perceptions cannot serve to prove their own veridicality. Instead, we are meant to scientifically investigate the properties of object surfaces, and then compare those properties to our ordinary perceptions. To do this, we may use advanced instruments to see how things look under extreme magnification, behave under highly controlled circumstances, interact with complex machinery, etc. This scientific investigation will allow us to discover ever more phenomenally determined properties, and — arguably — more explanatory and informative properties that incorporate endlessly more perspectives than those we evolved to naturally occupy. Still, none of this weakens

our phenomenal tether. For many, the fact that Cubey no longer looks reddishly under sufficient magnification — and perhaps, more importantly, that the way Cubey looks under magnification seems no more suited to reddishness than greenishness — shows us that the reddishness of our perception does not transparently reveal an aspect of Cubey. In reply, Kalderon sensibly argues for pluralism about colors (2007). Roughly, on such a view, phenomenal reddish perception is compatible with both the Martian's greenish perception and our microscope-aided perception because each reveals a different subset of Cubey's many properties. However, this pluralism is rarely defended with respect to spatial properties.

Papineau similarly describes examining the "worldly properties" — presumably through the techniques of the natural sciences — and comparing them to perceptual experiences of those properties:

> If the conscious character of sensory experiences did derive from the worldly properties they contain, then we would expect the conscious similarities between experiences to match the objective similarities between these Worldly properties. Yet it doesn't come out like that. The visual experience of blue is more similar to that of purple than green, even though in the physical world blue itself is closer to green than purple. (2021, 59-60)

Papineau presumes that similarities in how colored objects seem on an atomic scale can provide insight into "objective similarities", which can be used to assess the transparency of ordinary color perception. The fact that the objects seem one way under intense magnification might help us explain how the objects seem at larger scales, as accessing microscopic structures allows us to develop predictive scientific theories. Yet, facts about how the objects seem at a small scale do not alone entail that how the objects seem at larger scales is in contrast to their nature. After all, things that seem one way — with one

set of similarity patterns — under one set of conditions in the lab, actually do seem a different way — with a different set of similarity patterns — under different conditions. There can be no empirical justifications for assessments of which of the ways objects actually seem are ways that they should seem.[19]

For any property P that could be investigated empirically, no matter how scientific, there is a phenomenally determined property that is empirically indistinguishable from P. Cubey's phenomenal profile is not limited to simple appearance properties like looking reddishly from some arbitrary human perspective. We can scientifically investigate Cubey and learn about its atomic structure, chemical breakdown, inertial mass, or anything else, double-checking our calculations in as many ways as we can. Still, we can define the phenomenally determined property of seeming from every possible perspective like it would if it has this mass. Stipulating that Cubey has this phenomenally determined mass property stipulates that all parts of the world, such as our measuring instruments, scales, etc., seem as they would if Cubey had the relevant mass property. Then, just as we asked how our cubish perceptions can justify the claim that Cubey is not only cubical* but also noumenally cubical, we can ask: How can our empirical investigations justify the claim that Cubey has not only the relevant phenomenally determined properties but also the alleged noumenal properties? Scientific investigation does nothing to narrow the gap between the phenomenal and the noumenal or to demystify the noumenal. To narrow that gap, one needs to designate some perspectives as normatively authoritative so that how things seem from other perspectives can be judged in comparison.

[19] This general point is defended in Jones (1985).

If one wants to avoid divine normativity, one might turn to pragmatics for this justification. There is a notion that if we didn't perceive things how they actually are, then we would not be able to navigate the world, causing our species to die out. Perhaps, evolution guarantees, or at least makes likely, that our phenomenal experiences conform to the noumenal world to some extent. For instance, you might think that, at least with respect to spatial properties, dramatic phenomenal inversions, such as seeing cubes spherishly, could not occur without a dramatic drop in functioning, if at all. Papineau takes this alleged impossibility of phenomenal spatial inversion to show that "evolution has no doubt selected" the characters of our spatial experiences for representing corresponding worldly spatial structures (2002, 211). Thompson similarly argues that there must at least be "an isomorphism between relations within spatial experience and relations among external spatial properties" (2010, 176). And yet the arguments justifying these claims focus on the possibility, or lack thereof, of various inversions *within* the phenomenal realm. Generally, the considered inversion involves holding tactile perception fixed and attempting to alter vision. Yet, the claim that our actual visual phenomenology is well suited to our actual tactile phenomenology tells us nothing about how well suited any phenomenology is to "external spatial properties", unless we presuppose a privileged relation between the latter and our actual tactile perception. Perhaps we evolved to occupy certain combinations of perspectives because the information from those perspectives is comparatively easy to integrate. This says nothing of the relation between our perceptual phenomenology and the noumenal. Trying to ascertain that relationship is a non-starter without presupposing some perceptions transparent and veridical. Identifying a concrete obstacle to an inversion requires holding some perceptual phenomenology fixed to clash

with the inverted phenomenology. Trying to conceive of an inversion between our spatial perceptions and noumenal shapes amounts to nothing more than imagining that, as a matter of fact, cubes* are noumenally spherical and spheres* are noumenally cubical. By definition, such a world would be empirically indistinguishable from our actual world, given that noumenal facts entail only how the empirical facts ought to be assessed, not what the empirical facts are. Holding the phenomenally determined fixed holds the empirical world fixed. Neither we nor anything else could have evidence that we are in one world with respect to the noumenal rather than another. Our scientific mechanisms for predicting future evidence would be unchanged by this assumption of a noumenal inversion. Claiming that we will be somehow better off in one world rather than the other because of a correspondence between the noumenal and the phenomenal requires one to venture into anti-scientific mysticism and reject scientific explanations of physical events.

5. The Similarity Objection

So far, we've focused on perceptions with shared content that vary in phenomenal character. Yet, my view also implies the opposite possibility. Consider alien Al having a veridical perceptual experience that is identical — or at least, imperceptibly similar — in visual phenomenal character to your experience of Cubey. However, Al sees round, blue balls. Nothing in Al's visual field is red or cubical. If Transparency is true, the phenomenal character of a veridical perception is constituted by perceived aspects of the world. Here's a principle that seems to follow from Transparency:

> *Similarity:* Non-trivial similarities among the phenomenal characters of transparent, veridical perceptions should correspond to non-trivial similarities in the content of those perceptions.

Put simply: If my perception reveals what the world is like via phenomenal character, then any perception with similar character should say something similar about the world. (A "non-trivial similarity" is just a similarity that does not hold among all possible perceptions.) My view requires a counter-intuitive rejection of Similarity. Despite your visual phenomenal character being indistinguishable from Al's, your experience and Al's experience don't have any non-trivial shared content. Importantly, your perception and Al's perception *do* feature aspects in common, yet these are aspects that I have argued are shared by every perceptible thing. However, these trivial aspects are not all that you or Al see. You also see redness and cubicality (which, as I have argued, are identical to redness* and cubicality*), while Al may also see blueness and sphericality. Just as you can see a cube as a cube by seeing some parts of the cube, you can see the cubicality via some aspects of cubicality, which may overlap with the aspects of sphericality that Al sees. This might seem to contradict Transparency: An arbitrary quality, which bears no special relation to cubical things or spherical things, is standing in for cubicality in your perception and sphericality in Al's perception. Perhaps this means that cubicality and sphericality are only represented indirectly via arbitrary qualities rather than featuring in perception themselves.

To respond to this challenge, we must step back to evaluate what it takes for cubicality to feature in a perception. Even the most ardent Transparency theorist cannot insist that qualities of a single perceptual experience can constitute cubicality in its

entirety. Presuming you can perceive complex objects and their properties, it must be the case that you can perceive an object or property even if there is more to being that object or having that property than is captured by the character of your perception alone.[20] Consider a paradigmatic cubish perception. When you see Cubey on the table, you might see three sides of Cubey, even as you are said to see Cubey as a cube and being a cube requires having 6 sides. Now, consider fellow human Trish who lives in a world packed with "Trisided" objects that are composed of three thin square-shaped panels that are connected at right angles as if to form halves of hollow cubes. Trisides are extremely important to Trish's life, and she has therefore been accustomed to identifying them since infanthood. In contrast, cubic, solid objects have been rare in her environment. If Trish were in your shoes looking at Cubey, she would presumably erroneously see it as a Triside. Replacing Cubey with a Triside would have no effect on your visual experience; you would just erroneously perceive the Triside as a cube. There may be some phenomenal difference between seeing Cubey as a cube and seeing it as a Triside, just like there is a phenomenal difference between two ways of seeing the Necker Cube sketch. Regardless, there is an undeniable visual aspect in common between your experience and Trish's experience; you each perceive a physical aspect of the world that is part of both cubicality and Trisidehood. By perceiving Cubey as having this aspect, you perceive Cubey as a cube, even though there are other aspects of Cubey's cubicality that are not directly visually accessible to you. This is just like how you might see me as a human despite not directly seeing my heart or my

[20] For in-depth discussions of this issue and those in the vicinity, see: Thompson (1965), Davidson (1986), McDowell (1994), Byrne (2014), Siegel (2021).

brain or my toes. When you see o as P, your visual phenomenal character might feature only an aspect of P-hood that is shared by many non-P things.

If Transparency is right, then, for a perception to be Pish is simply for the perception to acquaint a subject with some property of the world, which we can call P†.[21] If Arbitrariness is true, all p† properties are trivial: if anything could have some property p†, every perceptible thing must be p†. Given that not every perceptible is red or cubical, redness† and cubicality† cannot be identical to redness and cubicality. If upon examining Cubey, *all* we perceived was that Cubey is red† and cubical†, we would fail to learn anything non-trivial about what Cubey is like. However, redness† and cubicality† are aspects of redness* and cubicality*. For instance, to be red* is, in part, to have redness† that is visible from the perspective of observers like us. Given your perspective, you are able to transparently perceive Cubey's redness* and cubicality* when you perceive Cubey's redness† and cubicality†. This is just like how you can perceive Cubey's cubicality by perceiving the spatial aspect that it shares with Trisides. Redness* and cubicality* are not trivial; the blue spheres Al perceives are not red* or cubical*. They have other properties, of which redness† and cubicality† are also aspects, although redness† and cubicality† will play different roles in those properties and be perceptible via different perspectives. Still, there is this similarity between your relation to Cubey's properties and Al's relation to the blue spheres; both feature reddish, cubish perceptions. Perhaps a futuristic neuroscience

[21] This distinction between pishness and P† does not imply two distinct qualities, one representing the other. If Transparency is right, then Pishness is simply the property of experiences which acquaints subjects with the P† quality. In Campbell's locution, if Transparency is true, then a perceptual experience is Pish iff it acquaints the subejct with P† ness. There is one quality P†; Pishness is just the property of the experiences featuring P†.

could characterize this similarity. With respect to what you learn about the world, your and Al's perceptions share only the trivial content of redness† and cubicality†.

6. The Metaphors of Perception

If we have no evidence for the noumenal, why is their assumption so widespread in the philosophy of perception? One culprit might be the historical legacy of supernatural grounds for the noumenal. Another potential factor is the ubiquity of a certain kind of metaphorical exercise that, I argue, begins with an implicit assumption of noumenality. In the philosophy of perception, metaphors abound. Is veridically perceiving like seeing shadows on a cave wall or like looking through a transparent window? A distorted window? Like watching a digital camera screen? Like reading a description? The metaphors of perception purport to describe how our perceptions of reality relate to what reality is like. In each metaphor, we begin with a set of objects, then describe a metaphorical experience of those objects: looking through a window at the objects, watching a live-stream of the objects, reading a description of the objects, etc. This experience is then compared to "the properties of the objects". In practice, this means comparing the metaphorical perceptual experience to an ordinary perceptual experience, which is presupposed as a normative standard for what the objects are really like.

The standard metaphor for transparent perception invokes a glass window on the world. Some allege that the complexity of visual processing refutes this model; the visual system seems to do more than passively allow the outside world in. To avoid this critique, Campbell provides a beefed-up window model involving a kind of "SmartGlass" that uses

information about its context to constantly adjust itself to remain transparent (2002, 119). With this "SmartGlass" model, Campbell demonstrates that there is no contradiction in supposing that perception is perfectly transparent while accepting that the perceptual system plays a substantial role in processing perceptual experience. It is an essential feature of SmartGlass that perceptual processing is not arbitrary, but instead occurs only as necessary to maintain perfect transparency. That is, the SmartGlass consistently alters itself to ensure that how the scene looks *sans* glass is identical to how the scene looks through the glass. How the scene looks *sans* glass thus dictates which perceptions count as transparent and veridical. If Campbell's model is correct, we are always behind the metaphorical SmartGlass. All possible perceptions — including those of our scientists as they peer into microscopes, read brain scans, etc. — tell us how reality seems from behind the SmartGlass. Other possible perceivers presumably look at the world through different metaphorical mediums with varying degrees of transparency. The facts about how the world looks *sans* glass are not phenomenally determined. All empirical evidence is about how things look through a metaphorical medium. The facts about how the world looks *sans* medium only interact with the empirical facts insofar as they entail what the empirical facts *should* be by grounding the facts about whether the SmartGlass is effectively remaining transparent. For this metaphor — and its corresponding Transparency-Arbitrariness tradeoff — to accurately represent perception, there must be noumenal facts occupying the role of the *sans* medium facts.

Similarly, the central anti-Transparency metaphor compares perceiving to reading linguistic descriptions. On this view, the characters of perceptual experiences are like

symbols in a language developed by your brain to track worldly facts. Papineau most notably defends this model (even while ultimately rejecting full Arbitrariness):

> [C]onscious sensory properties are very much like written marks on paper. In themselves, they are dumb arrangements of contentless signs. While they do represent worldly facts, this is only in virtue of further contingent facts about the way they are embedded in the wider world. (2021, 5)

Even if such a description is accurate, it is opaque; the qualities of the description, i.e. the shape or sound of the letters, the colors or pitch of the words, etc., are not the qualities of the described world. For instance, the qualities of 'The sky is blue" are not the qualities of the blue sky. While the experience of looking through a perfectly transparent window perfectly resembles the experience of looking at the world *sans* window, the experience of reading a description of the world has no essential similarities to the experience of direct perception of that world. The opacity of language is also intimately related to its arbitrariness; any set of symbols could, in some language, describe the blue sky because the qualities of the symbols need not conform to the qualities of the sky. The description is opaque because reading the description does not resemble perceiving the described world in the way that seeing the world through the window resembles standardly perceiving the described world.

In both cases, facts about how the world seems in ordinary perception play the role of the noumenal. They are not metaphorically phenomenally determined given that they are not determined by the metaphorical phenomenal facts (the facts about how things look through mediums or how descriptions of things look). These ordinary perceptual facts instead only serve as the standard of comparison used to determine the veridicality and transparency of metaphorical perceptions. Thus, for any of these metaphors to be

illuminating, the ordinary perceptual facts must themselves be assessable relative to some further, phenomenally undetermined facts, just as the metaphorical perceptual facts are assessable relative to the ordinary perceptual facts. Once you accept the task of trying to describe a metaphorical stand-in for perception while continuing to accept ordinary perception as the standard-bearer for transparency and veridicality, you have already accepted Incompatibility and postulated the noumenal.

The persistence of talk of the "veil" of perception similarly promulgates the noumenality myth by implying some unveiled perception. Tye accuses Shoemaker's phenomenal property view of colors of "effectively draw[ing] a veil over the colors" and thus "erecting an appearance/reality distinction for the colors themselves" (2000, 464). Shoemaker rebuts this criticism:

> But to say that there is such an appearance-reality distinction suggests that there is a unique way a color would look if one were perceiving it "as it is," and that in the case just described at least one of the perceivers would be misperceiving the color that they perceive differently. And of course my account denies this. Insofar as colors are nonrelational properties of objects, there is no way a color looks simpliciter; there are only the ways it looks to observers with visual systems of certain sorts. (2000, 466)

What Shoemaker says about colors is roughly what I say about all perceptible properties: there is no way things look — or more generally perceptually seem — simpliciter. In saying that our perceptions do not conform to some universal standard of how things ought to seem, I do not deny that we perceive things how they are; I merely deny that there is a way things are which includes one universal standard of how things ought to seem.

7. Conclusion

I advocate abandoning the noumenal — or, at least, taking seriously the project of its defense. The phenomenal world contains not only all properties of which we have evidence but all properties for which anyone could have evidence. Our scientific worldview offers no justification for believing in imperceptible souls of objects silently assessing how things empirically present themselves, yet incapable of affecting that presentation. Giving up the noumenal, we can fully accept the insight that the character of our perception depends on arbitrary accidents without denying that perception transparently reveals what the world is like.

Perhaps this conclusion should humble us insofar as we like to think that our perceptions carve privileged joints of reality. Yet we must not falsely equate informative, veridical perceptions with omniscient, ideal perceptions. After emphasizing the extent to which our perceptual systems developed to access aspects of the world that are evolutionarily important to our fitness, Hoffman ultimately concludes that the perceived world is unreal and cannot exist unperceived. This reasoning relies on the false Transparency-Arbitrariness tradeoff. Our perceptions can transparently reveal true aspects of the world, even if *which* aspects feature in our perceptions depends on arbitrary facts about us. Accepting that we cannot access some ideal, noumenal realm does not require giving up on perceiver-independent properties. Phenomenally determined properties entail how bits of the world will seem to someone like you — and thus, they can feature in your evidence — but they do not depend on any perceiver's existence and they are not perceiver-relative. Phenomenally determined properties are real, objective properties of

the physical world. They do not depend on anyone perceiving them. Claims about them could never be exhaustively translated into claims about our evidence.

Through science, we continue to discover more complex and explanatory phenomenally determined properties, even if we are fundamentally constrained by our particular perceptual capacities. Rejecting the noumenal does not lead to Hoffman-style anti-realism. For comparison, imagine that Lewisian Modal Realism or some Multiverse scenario is true such that all of your empirical investigations consist of investigations into which world you are in, as opposed to investigations into the one true world. Learning this might humble you, but it would not and should not lead you to conclude that your perceptions of your world are less transparent, even if you conclude that they are less important. Similarly, the claim that there are infinitely many aspects of the world, of which you perceive a subset determined by your evolutionary past, does not endanger the thesis that your perceptions transparently reveal (some of) reality.

References

Brewer, Bill. 2007. “Perception and Its Objects.” *Philosophical Studies* 132: 87–97.

Byrne, Alex. 2001. “Intentionalism Defended.” *The Philosophical Review* 110: 199–240.

——2014. “Perception and Evidence.” *Philosophical Studies* 170 (1): 101–13.

Campbell, John. 2002. *Reference and Consciousness*. Oxford University Press.

Egan, Andy. 2006. “Appearance Properties?” *Noûs* 40: 495–521.

Cumpa, Javier, and Bill Brewer. 2019. *The Nature of Ordinary Objects*. Cambridge University Press.

Davidson, Donald, and Ernest LePore. 1986. “A Coherence Theory of Truth and Knowledge.” *Epistemology: An Anthology*, 124–33.

Descartes, Ren. 1681/1984. “Meditations. V: Cottingham, J., Stoothoff, R.” Murdoch, D.(ur.), *The Philosophical Writings of Descartes*, Cambridge University Press.

Egan, Andy. 2006. “Appearance Properties?” *Noûs* 40 (3): 495–521.

Gosselin, Frédéric, Mégan Brien, Justine Mathieu, and Ariane Tremblay. 2022. “A Collection of Stationary Objects Flashed Periodically Produce Depth Perception under Ordinary Viewing Conditions.” *Journal of Vision* 22 (14): 4134.

Harman, Gilbert. 1990. “The Intrinsic Quality of Experience.” *Philosophical Perspectives* 4: 31–52.

Hoffman, Donald D. 2019*. The Case Against Reality: How Evolution Hid the Truth from Our Eyes*. Penguin.

Jones, O. R. 1985. “The Way Things Look and the Way Things Are.” *Mind* 94: 108–10.

Kalderon, Mark Eli. 2007. “Color Pluralism.” *The Philosophical Review* 116: 563–601.

Logue, Heather. 2012. “Why Naïve Realism?” *Proceedings of the Aristotelian Society* 112: 211–37.

McDowell, John. 1994. “The Content of Perceptual Experience.” *The Philosophical Quarterly*. 190–205.

Mendelovici, Angela. 2013. “Reliable Misrepresentation and Tracking Theories of Mental Representation.” *Philosophical Studies* 165: 421–43.

Papineau, David. 2021. *The Metaphysics of Sensory Experience*. Oxford University Press.

Pautz, Adam. 2021. *Perception*. Routledge.

Russell, Bertrand. 1912. *The Problems of Philosophy*. Oxford University Press.

Searle, John R. 1980. “Minds, Brains, and Programs.” *Behavioral and Brain Sciences* 3: 417–24.

Shoemaker, Sydney. 1994. “Phenomenal Character.” *Noûs* 28: 21–38.

——2000. “Phenomenal Character Revisited.” *Philosophy and Phenomenological Research* 60: 465-7.

Siegel, Susanna. 2011. *The Contents of Visual Experience*. Oxford University Press.

——2021. "The Contents of Perception", *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.)

Sellars, W. 1962 “Philosophy and the Scientific Image of Man”. *In the Space of Reasons.* Selected.

Thompson, Brad. 2010. “The Spatial Content of Experience.” *Philosophy and Phenomenological Research* 81: 146–84.

Tye, Michael. 2000. “Shoemaker’s The First-Person Perspective and Other Essays.” *Philosophy and Phenomenological Research*: 461–64.

——2003. “A Theory of Phenomenal Concepts.” *Royal Institute of Philosophy Supplements*: 91–105.

Chapter 3: What Mary Already Knew

ABSTRACT: This chapter defends an empiricist constraint on understanding language. I argue that anyone who understands language must have some "Ostensive Competency." That is, for at least some linguistic expressions, one must properly associate e with a perceptual representation of a part of the world described by those expressions. I argue that the arbitrariness of language prevents anyone — regardless of intelligence, data, etc. — from learning the meaning of words merely from learning how words relate to other words or other arbitrary symbols. If my view is right, then any knowledge of the world necessarily relies on a foundation of knowledge about how the world perceptually seems to the knower.

1. Introduction

In 1982, Jackson introduced the now-famous Mary, a color scientist raised from birth in a black-and-white room. Inside her room, Mary learns all "physical facts" about color. Still, Jackson claims, Mary must learn something new upon finally seeing colors. She will learn, for example, what it is like to see red. If this is right, then not all facts are physical facts — or so argues Jackson.

Debate about Mary centers the peculiar piece of information — a "phenomenal fact" about what it is like to see red — that Mary allegedly learns upon leaving the room. Many accounts are offered about how this "phenomenal fact" relates to the "physical facts" that Mary learns inside the room. My concern here is how Mary learns those physical facts. Take the fact that red has a dominant wavelength of approximately 625–740 nanometers or that strawberries are red. It is generally taken for granted that Mary can learn these facts inside the room, with minimal discussion of how this learning occurs. Some argue that Mary doesn't have our ordinary concept red, a concept that you can only acquire via certain

phenomenal experiences. If that is right then Mary (and the blind) cannot know that strawberries are red, but presumably can know instead that strawberries are red*, where 'red*' expresses a different concept of red, one that can be acquired without having certain phenomenal experiences. Regardless of such complications, the physical facts can be easily taught to Mary inside the room, with no particular phenomenal experience required.

Of course, this learning of the physical facts relies on *some* perceptual experiences — seeing a textbook page, hearing a lecture, etc. Yet, in contrast with the special phenomenal fact about what it is like to see red, learning these physical facts doesn't seem to require any particular experiences. Mary could learn them by reading black-and-white textbooks, or by watching black-and-white videos, or by hearing podcasts. The textbooks could be printed in Times or Calibri. They could be written in English or (if Mary speaks Korean) in Korean; if Mary reads Braille she could learn the physical facts by touch.[22]

The physical facts, then, can be represented via *arbitrary* symbols. Any symbols that could represent any fact could, in the proper context, represent these physical facts. However, a fact can only be taught via arbitrary symbols if the speaker already understands those symbols. For instance, when Mary reads a textbook passage including 'Red has a dominant wavelength of approximately 625–740 nanometers', she only learns a new fact about redness because she understands these English words. The complete process of teaching her those facts thus includes teaching her the meaning of those words.

[22] Nagel (1974) helpfully distinguishes between "subjective" phenomenal facts — like what it's like to be a bat — which essentially depend on particular viewpoints and more "objective" facts which can be accessed from more varied viewpoints.

It is overly simplistic to say that we can teach Mary the physical facts through language alone; the learning of that language is an essential part of the learning of the fact.

This chapter considers what it takes learn that language. I will argue that, when Mary learned English, this necessarily involved acquiring some *ostensive knowledge* connecting words and their perceptually presented referents. Paradigm cases of ostensive knowledge (in the case of learning count nouns) involve Teacher demonstrating a perceived F, saying '*That* is called 'an F'" (or, more simply, '*That* is an F'). The Learner thus learns that the word 'F' applies to this particular object (which is in fact an F), under a certain perceptual mode of presentation, or as represented perceptually. For example, Teacher points to a nearby cat, which Learner can see, and says '*That* is a cat'. Here vision is crucial to what is learned, or so I will assume: Learner could not learn the crucial fact conveyed by '*That* is a cat' via some other sense modality. (I have only given some paradigm examples of ostensive knowledge, not a fully general characterization; I will attempt to fill out the notion in more detail later.)

There is no *particular* piece of ostensive knowledge that Mary must have learned, although I argue that some ostensive knowledge is a necessary foundational component of Mary's ability to understand the language required to learn the physical facts. This claim conforms to the classic, Lockean Empiricist view of language, as endorsed by Russell in 1948:

> All nominal definitions, if pushed back far enough, must lead ultimately to terms having only ostensive definitions, and in the case of an empirical science the empirical terms must depend upon terms of which the ostensive definition is given in perception. (1948, 248)

Throughout the middle of the 20th century, this kind of Empiricist wisdom was stretched to its extreme, as "Logical Empiricists" endorsed Verificationist principles essentially requiring all knowledge reduced to kind of ostensive knowledge. After this project was generally understood to have imploded in inconsistency, the philosophical spirit overcorrected (or so I argue) by giving up on the essential Empiricist insight that that our ability to represent the world linguistically (or otherwise) necessarily stems from our ability to represent the world perceptually.

Accepting my empiricist thesis does not automatically resolve Jackson's puzzle of how the phenomenal facts relate to the physical facts, but hopefully renders the situation less mysterious. According to the picture I defend, Mary's in-room learning necessarily includes a foundation of knowledge of the form *This is an F*, with 'this' receiving its content from perception.[23] This foundational learning necessarily relies on perceptual experience in just the same way as the fact that Mary learns upon leaving the room. Mary's understanding of the physical facts, representable via arbitrary symbols, thus cannot be divorced from her knowledge about what it is like to perceive the world. While learning physical facts representable via arbitrary symbols requires *some* amount of perception-mediated knowledge, no particular perception-mediated fact is required. Thus, before leaving the room, Mary already knew some, but not all, facts about what it is like to perceive the world, and this incomplete perceptual knowledge sufficed as a foundation for

[23] Given that my concern is the relation between perception and language, I will focus on the form of knowledge with respect to perceptual states, rather than "phenomenal states", to avoid passing judgment on the possibility of a "zombie perceiver" of the sort discussed in Chalmers (1996).

her ability to learn any fact that could be represented via arbitrary symbols. Upon leaving the room, she simply learns one more fact about what it is like perceive a bit of the world.

2. How to Learn a Language

We can sort language learning into three rough buckets: "The Dictionary Method", "Immersion Training", and "Ostensive Learning". While rather rare in the wild, the Dictionary Method serves as the paradigm of language learning. One way to learn the meaning of a new word is to look it up in a dictionary or be told its definition. For instance, one could learn the meaning of 'bachelor' by seeing that it means 'unmarried man', given that one already understands 'unmarried man'. A Spanish speaker could learn 'bachelor' with a bilingual dictionary that translates 'bachelor' to 'soltero'. In its idealized form, the Dictionary Method works when you learn the meaning of an unfamiliar word by learning that the meaning of the word can be expressed in a bit of language that you already understand. This is undoubtedly one way in which someone can learn a new word. Yet, even setting aside Quinean qualms about synonymy and translation,[24] it is often pointed out that most dictionary entries do not contain definitions, in the minimal sense of necessary and sufficient conditions for the word to apply. Dictionaries instead typically provide rough approximations to definitions, supplemented with examples of how the word is used and/or contingent facts about the word's referent. For example, the Merriam-Webster Definition of 'celebrate' describes some of the activities that might be classified as celebrations — i.e. "to honor (an occasion, such as a holiday) especially by solemn

[24] Given that this chapter is about the requirements for achieving *any* understanding of word meaning, whether or not any definition or synonym is "perfect" is irrelevant, given that one can certainly achieve some understanding of word meaning from imperfect definition or synonym.

ceremonies or by refraining from ordinary business", "to mark (something, such as an anniversary) by festivities or other deviation from routine" — as well as some examples of typical phrases containing 'celebrate' — i.e. "The nation celebrates Memorial Day.", "celebrated their 25th anniversary"). This sort of "definition" tells us what typical celebrations might be like rather than providing precise conditions for application of 'celebration'. Then, Merriam-Webster provides an alternative set of "kid's definitions", such as "to observe in some special way (as by merrymaking or by staying away from work". This definition provides even less precise information about 'celebration', but presumably is designed to help kids to start to understand some ways the word might be used. Much dictionary-moderated language learning thus fits more neatly into our second bucket: Immersion Training.

Most human language learning does not occur through explicit lessons in word meaning. When you immerse humans, especially young ones, in linguistically rich environments, they come to "pick up" the meanings of words, quickly acquiring language usage patterns from their environment. While the details of this process are a subject of persisting study and controversy, we know that the speech that we hear and texts we read provide evidence of the meanings of words. When we speak, the reactions of others also help train us in proper usage. Say you have never encountered the word 'apricating' before hearing the following conversation:

A: I spent my vacation apricating by the lake.
B: That sounds wonderful.

Even without a "definition", you have learned quite a lot about the meaning of "apricating", provided you already understood the other words used. You know that apricating is a pleasant activity sometimes carried out by lakes. As you hear 'apricating' used in more contexts, you may automatically adopt it as part of your own vocabulary. This process happens constantly — mostly without explicit metalinguistic reasoning — as we are exposed to novel vocabulary in novel contexts. I use 'Immersion Training' specifically to describe this phenomenon of learning a word by hearing (or reading) how that word is used in relation to other words. However, this is not the only way to learn language from one's environment. We also can learn about meaning by witnessing how word use interacts with the wider material world. For example, you might hear someone say, 'Look at Sue apricating by the lake' while gesturing towards Sue. Then you see an example of apricating for yourself. Learning like this, while often spurred by a kind of linguistic immersion, belongs better in our third bucket: "Ostensive Learning".

Ostensive Learning occurs when someone learns a new word by perceiving a bit of the world described by that word. If you ask me what 'yo-yo' means, instead of describing the meaning in words, I might respond by *showing* you a yo-yo, saying "*this* is a yo-yo." Now, you can see and feel an example of the sort of thing described by 'yo-yo'. This sort of lesson has often been called, somewhat misleadingly, "Ostensive Definition", although as Wittgenstein argued through his criticisms of the "Augustinian Conception of Language", no complete word meaning — or, "definition" as it is usually understood — can be fully, unambiguously conveyed by pointing to a thing and saying or implying that the word refers

to the indicated thing.[25] Yet even Wittgenstein accepted the role of Ostensive Learning in language learning:

> An important part of the training will consist in the teacher's pointing to the objects, directing the child's attention to them, and at the same time uttering a word; for instance, the word 'slab' as he points to that shape. (I do not want to call this "ostensive definition", because the child cannot as yet ask what the name is. I will call it "ostensive teaching of words".—I say that it will form an important part of the training, because it is so with human beings; not because it could not be imagined otherwise.) This ostensive teaching of words can be said to establish an association between the word and the thing. (1953, 4)

I'll categorize any learning as Ostensive Learning that allows one to associate language with a perceptual state/experience of a bit of the world described by that language. This association need not — and usually won't — involve any metalinguistic representation. For instance, if a toddler calls 'Mama' because they want their mother, whom they visually recognize as their mother (and the object of their demand), this is a way to correctly connect 'Mama' to their perceptions of their mother. You might object to my characterization of this association as linking 'Mama' to *perceptions* of the mother rather than simply "the mother". (After all, Wittgenstein refers to "an association between the word and *the thing*", not a word and a perception of the thing.) I don't deny that the child links 'Mama' to *their mother*, yet I will insist that, insofar as the learning is "Ostensive Learning", the link between the word and the perceptual representation must be present.[26]

---

[25] You might think that teaching you that *this particular yo-yo* is called 'Ya-Ya' is more straightforward. However, Wittgenstein famously showed that even this situation is more complicated than it seems. To *fully* understand this proper name, it seems I must already understand the concept of a yo-yo. However, I will leave these issues to the side, given my focus on merely necessary conditions for understanding and my focus on how one attains any understanding, as opposed to complete understanding, of word meaning.

[26] You might theorize that this link should be understood as occurring *via* some concept of the mother, such that 'Mama' and the perceptual representations of the mother are connected only insofar as they both are associated with the mother concept. That understanding is fine by me.

For instance, say that I've been told in the past about Dan's estranged sister. Now, Dan tells me that his sister is called 'Liz'. I newly associate 'Liz' with Dan's sister. Yet, this is not Ostensive Learning. Instead (à la the Dictionary Method) I've learned to associate a new term, 'Liz' with a referent represented via another term, as opposed to a perceptually represented referent. Similarly, if you present me with a steel box, and explain to me that the hidden object inside is called a 'Whyz', I learn that 'Whyz' refers to the object which you have verbally described to me as being in the box. However despite the prop and despite the sense in which you have "presented" me with the Whyz, I've done no Ostensive Learning of 'Whyz' because I have not perceived the Whyz myself.

Ostensive Learning can happen via explicit instruction resembling the definition-providing Dictionary Method (e.g., Teacher points, saying, '*This* is a cube'.) or more implicit, Immersion-Training-esqe absorption of how words are used in relation to perceptible aspects of the environment (i.e. people often say 'Bob' when addressing Bob). Either way, Ostensive Learning involves associating new language with the learner's own perceptual representation of its referent,[27] while the Dictionary Method and Immersion Training teach new language only via other language.

This chapter argues that Ostensive Learning — or, at least, the association between perceptual representations and words that Ostensive Learning facilitates — is a necessary

[27] I use 'referent' as a shorthand for 'whatever is described by the language'. Yet, in addition to nouns, one can also learn verbs (like 'apricating' in the case above), adjectives, and adverbs this way. One might learn 'round' by being shown the round things, 'walking' by being shown walking things, etc. Any language that might properly show up in a description of certain perceptual content could be learned in this way, as one is taught to perceptual experiences featuring that content with the relevant language. (Again, "learning a word", in my sense, need not entail attaining any kind of complete understanding of word meaning.)

component of language learning, contrary to Wittgenstein's parenthetical remark. For much of philosophical history through the middle of the 20th century, as Lockean Empiricism reigned, the foundational role of Ostensive Learning hardly needed defense. In 1956, Whiteley offers the following as a bit of undefended ground-clearing:

> Let us also assume that there is such a procedure as Ostensive Definition, and that it is the means whereby men learn the meanings of most, if not all, of those elementary expressions in their languages in terms of which other expressions are defined. (332)

Here is a simple argument for the claim that Ostensive Learning is necessary for language learning:

> P1. The available methods for language learning are learning words from other words and learning words Ostensively.
>
> P2. One cannot learn a new word from other words if one does not understand those other words.
>
> P3. One cannot understand words one has not yet learned.
>
> C: All language learners must learn some words Ostensively.

Davidson's famed "Swampman" example offers an immediate objection to P3: Take a typical adult man fluent in English. Now, in a complete physical coincidence, a perfect, molecule-for-molecule duplicate of that man comes into existence out of a swamp. Many philosophers say that Swampman understands English upon emergence, contradicting P3. Furthermore, just after his emergence, we could presumably teach Swampman some German by offering German-to-English translations, even though Swampman never learned English. He would then be a language learner who never learned words Ostensively. To

avoid such issues, we can focus on the outcome of Ostensive Learning rather than the process itself. As discussed above, Ostensive Learners come to appropriately associate language with perceptual representations of the world; we can call such associations "Ostensive Competency" to allow for the possibility of Swapman-like creatures who (by innate luck) have the sorts of perceptual-linguistic associations acquired during Ostensive Learning.[28]

I thus defend the thesis — which I'll call *Perceptual Foundationalism (PF)* — that understanding language requires some Ostensive Competency. PF does *not* entail that there is any particular bit of language that must be understood Ostensively. Instead, PF offers the holistic claim that anyone (human or otherwise) who understands language (of the sort that can represent contingent states of the world) must have Ostensive Competency with respect to *some* words. Then, other words might be understood in terms of Ostensively understood words.

If PF is correct, then Ostensive Teaching will be a necessary component of teaching Mary the language required to teach her the physical facts. Section 3 defends this result by rejecting the possibility that Immersion Training alone could suffice for language learning. Section 4 similarly rejects the possibility that we might learn language via translation into a "Language of Thought" without relying on perception. Ultimately, I argue that one cannot come to understand a language merely by learning how that language interacts with other arbitrary representations, linguistic or otherwise.

[28] While this "Ostensive Competency" might generally come in the form of Ostensive Knowledge, plausibly Swampman, ostensively competent by luck, lacks the relevant knowledge.

3. Extreme Immersion Training

It is uncontroversial that perception is required for Mary to learn a natural language, for instance English. If Mary has no perceptual capacities, she cannot become aware of English sentences and therefore cannot represent them. To learn English, she must perceptually represent some English words. However, is perception of language enough? I argue that the answer is "no". For at least some bits of language, Mary must perceptually represent bits of the world described by that language and then learn to associate these perceptions with the language. Jackson introduces Mary as "a brilliant scientist who is, for whatever reason, forced to investigate the world from a black and white room *via* a black and white television monitor" (130). As a parenthetical response to anticipated skepticism, Jackson adds that "[i]t can hardly be denied that it is in principle possible to obtain all this physical information from black and white television, otherwise the Open University would of necessity need to use colour television" (130). However, the story cannot begin with Mary watching advanced science lectures from the Open University. First, she must develop the English competency required to understand such lectures. The example requires that Mary has been trapped in the black and white room since she was an illiterate, ignorant baby. Despite her color deprived surroundings, Mary is a typical, if abnormally intelligent, baby. The task of her teaching her English is thus not much different than the task of teaching any child English, which is presumably why not much attention has been given to this part of her education.

Imagine that you are the caretaker trying to teach Baby Mary English.[29] Presumably, you will engage in the standard practice of repeatedly using words while showing Mary the things named by those words. For instance, you will show Mary the block, let Mary touch the block, while saying 'block' (or 'Here's a block'. etc.). Over time, she learns, perhaps, to answer 'block' when you ask, 'What is this?' and point at the block, and to say 'block!' when she wants you to hand her the block. Mary thereby correctly associates 'block' with certain perceptions of the block. To see the importance of her perceptions of the block to her understanding of 'block', consider that if the block looks radically different from underneath, and Mary has never seen the block from this angle, she will not know to apply 'block' when she visually represents the block from underneath. To learn that 'block' refers to the block, she must be able to represent the block independently of 'block'. To accept that she has this association between 'block' and her perception of the block, one need not imagine her thinking metalinguistic thoughts; it is enough that, upon having a certain sort of perception of the block, she knows that *that* [the thing she visually perceives] is a *block* [which she can express using 'block']. You will presumably teach Mary many words in this way as she begins to acquire language. Also, Mary can learn words this way even without explicit teaching; that is, she will hear a word in conjunction with perceiving a thing in such a way that she comes to recognize the perceived thing by its name. Then, once she has some language, her vocabulary will quickly balloon to include many words for things and situations that she cannot or does not perceive.

[29] So long as we are sticking to the case in which Mary is a relatively normal, helpless human baby, we ought to include a caretaker for Mary in her room, rather than just a television. We could either alter the appearance of the caretaker or alter Mary's eyes to maintain the black-and-white facade.

The story of Helen Keller offers some reason to accept that this Ostensive Leaning is a necessary part of language learning. In 1887, Annie Sullivan, later deified on stage as "The Miracle Worker", succeeded in teaching English to her deaf and blind student, Helen Keller. Despite Keller's lack of vision and hearing, Keller's early word acquisition was structurally just like the fictional, typical story above of Mary learning to associate 'block' with her perceptions of the block. Substituting touch for vision, Sullivan repeatedly spelled out the names of objects into Keller's skin, while allowing Keller to tactically perceive the named objects. Then, famously, one day, when Sullivan put Keller's hand under a water pump while spelling out the word water into her palm, Keller finally came to understand the meaning of 'water'. Here's how Keller herself later described this breakthrough moment: "[S]omehow the mystery of language was revealed to me. I knew then that "w-a-t-e-r" meant that wonderful cool something that was flowing over my hand"(23). Keller thus describes finally coming to associate 'water' as it is spelled into her skin with the water as she feels it on her hand, just as Mary might learn to associate 'block' as it is spoken to her with the block she sees and feels. Eventually Keller developed a vast vocabulary, becoming an eloquent writer. Still, in the story of Helen Keller, it is this moment, in which Keller connects a word to a perception of a part of the world, that is celebrated as the incredible breakthrough that made it possible for her to understand language.[30]

Although this may demonstrate that it would be quite difficult for a human to learn language without Ostensive Learning, it does not prove that it is impossible. Return to you

[30] The relevance of the tactile perception of the water also counteracts the idea — endorsed in Rapaport (2006) and refuted in Ford (2011) — that Keller should be understood as enclosed in a Searlean "Chinese Room", with access only to syntactic properties of language.

teaching Mary English. Let's say you just talk to her in English, without any use of ostension, pointing, holding props, etc. You reward her for making noises that approximate appropriate English responses, perhaps starting with extremely low standards for responses and then raising the standards if/as she improves. In doing this, you might train her to utter some appropriate noises at appropriate times. Like a trained monkey, she will eventually say 'Mary' in response to "Who are you?', as well as answer some other questions correctly. Still, it is doubtful that she will understand *any* of what is being said. Her situation resembles that of Searle in his famous "Chinese Room": he knows how to associate symbols (verbal symbols, in her case), but she doesn't apparently need to understand any of their meanings (Searle 1980).

Here is an argument for the claim that Mary cannot understand the words she hears or responds to: The meanings of 'Who are you?' and 'Mary' are completely arbitrary, even holding fixed that the latter is the correct answer to the former in Mary's context. These phrases could mean just about anything. There is a possible language just as good as English in which these expressions translate to the English phrases 'Who am I?' and 'teacher'. There's still other equally good languages in which these phrases translate to 'What color is the sky?' and 'blue', 'Where are we?' and 'Boston', or 'What are you sitting on?' and 'a chair'. We might even imagine that you, Mary's teacher, are familiar with some of these alternative languages. Now we can ask what determines that you taught Mary English rather than some alternative language in which 'Who are you?' is a question about the sky. The only answer we could give is that this determination depends on your intention to teach one language rather than another. Yet, this response gives rise to difficulties. What if you lack any particular intention or what if your intention shifts during

the process? What if, instead of a live teacher, we employ a smart robot to teach Mary these expressions?

We could imagine a set of qualitatively identical scenes, each with a teacher teaching Mary (or one of her "twins"), such that each teacher belongs to a society with a different version of English. If Mary understands the English question 'Who are you?', then her twins, all with mental states qualitatively identical to hers, must similarly understand the questions asked to them via 'Who are you?'. Similarly, when Mary understands 'Mary', her qualitative mental state must be compatible with understanding a term that refers to chairs, colors, places, shapes, actions, etc. Even Putnam, who invented "Twin Earth" to argue that meanings "just ain't in the head" didn't think that mental states were *irrelevant* to meaning (1975, 227). Following Putnam, we may accept that meaning may be often partially determined by something outside the mind, including the intentions of the person who taught you a new word. For instance, let's say you tell me, "I am worried about my sick friend Kate." As it happens, you have two friends called 'Kate', both of whom are sick and worthy of your worry. Yet, in the moment of your utterance, you are thinking of 'Kate A.' Then, I go on to ask God to help Kate. It seems plausible that when I use 'Kate' in my prayers, I successfully refer to Kate A. because of your intention when you spoke to me. In this case, I use 'Kate' with the intention to refer to the friend that you told me about, and since your intentions when speaking influence the content of your divulgence, those intentions can play a role in determining my intentions.[31]

[31] We might even imagine that I don't know if Kate is a person or a dog or car; Still, I know that Kate is a particular thing that you are worried about. I am in a position to use 'Kate' meaningfully in conjunction with words that I already understand.

In contrast, in our scenario, literally all Mary knows about 'Mary' is that it is the rewarded answer to 'Who are you?'. All she knows about 'Who are you?' is that its answer is 'Mary'. If one takes a holistic view, according to which these circular definitions alone are enough for attaining some understanding of these terms, then one must be committed to the view that qualitative mental states do nothing to constrain mental content. At this point, a holist may object that I have rigged the game by providing a case in which Mary learns so pathetically little, so as to understand practically nothing about the meaning of these words. Maybe her lack of understanding is due to the tiny scale of her education as opposed to the absence of Ostensive Learning. So let's imagine that we keep instructing Mary in this fashion, teaching her more and more complex sentences, training her to engage in ever more complex conversation. She moves well beyond rote memorization, eventually consistently producing novel, extended, appropriate responses to novel comments and questions.

In the past, we may have thought this scenario not merely unrealistic but impossible. However, in the age of Large Language Models, we must admit that, so long as we allow Mary sufficient (plausibly, super-human) intelligence and time, it is in principle possible to develop advanced language abilities from merely reading texts in a novel language, trying to respond to the text, and receiving feedback on your responses. So now we have SuperMary. The structure of her learning is like that of the Mary we have been describing. She is read bits of English and rewarded based on her responses to the texts she hears. However, SuperMary is a supergenius with absurdly large memory who has been training for a millennium. She therefore can give long, eloquent answers to complex questions and engage in sophisticated dialogue that is indistinguishable from other

intelligent English speakers. SuperMary does not merely know that 'chair' is the answer to 'What are you sitting on?'. She can employ 'chair' correctly in indefinitely many novel scenarios, answer all kinds of novel questions about chairs, etc. Unlike Searle in the Chinese Room, Mary has not simply memorized massive numbers of rote replies; she uses her dynamic language skills to respond appropriately to novel statements that are not stored in her memory. Many will insist that SuperMary must understand 'chair'.

Yet the arbitrariness of language persists, even as Mary learns *more* language. Take all of the linguistic data that SuperMary — or an LLM — receives, both through passages read/input and feedback received. SuperBari, a duplicate of SuperMary, has been sent to an alien planet to be raised in her own black-and-white-room. Is there a possible language—call it *Alienese*—such that SuperBari, learning Alienese, receives all of the same linguistic data as SuperMary, yet the meanings of the words of Alienese differ dramatically from the meanings of their English counterparts? For instance, in Alienese, 'chair' might mean *pineapple*, etc. Stipulating that Alienese must be just as learnable and not any weirder or more unnatural than English, structural constraints would plausibly prevent us from defining Alienese by merely mapping words of Alienese to meanings expressed in English such that each passage in our data "translated" from Alienese to English produced a new set of English passages fit to replace the original English data. However, we do not need to restrict ourselves to meanings easily expressed in English. Perhaps speakers of Alienese (Aliens) talk about all kinds of alien things from their alien planet for which we have no name. Perhaps they differ radically from us in their perceptual capacities, interests, motivations, environments, etc.

With these constraints of sameness between English speakers and Aliens thus lifted, the task of defining the words of Alienese such that SuperBari can receive the same data as SuperMary — despite dramatic differences in meaning between the English words and the Alienese words — suddenly seems possible, while arduous. If we want to take as a starting point that 'chair' in Alienese refers to pineapples, it seem that we could, in principle, define the rest of the terms in Alienese so that the Alienese data makes just as much sense relative to the Aliens' lives as the English data makes sense relative to human lives.

If this is indeed possible, then we return to the dilemma faced above: How can we say that SuperMary understands the meaning of 'chair' when a perfect duplicate of her might be using 'chair' to mean practically anything? Again, we can ask, what if SuperMary is taught by a robot or merely by written text? Can it be the case that SuperMary understands the meaning of 'chair', even though the source of teaching leaves it is indeterminate whether 'chair' refers to chairs or pineapples? Or can it be the case that SuperMary and SuperBari, perfect duplicates, both understand the meaning of 'chair', but due to the intentions of their masters, SuperMary understands 'chair' to refer to chairs and SuperBari understands 'chair' to refer to pineapples? What if they swap rooms half-way through their learning: do they still understand the meaning of 'chair'?

If you accept the described possibility, then you must accept that the place that a term holds in a linguistic structure is ultimately just as arbitrary with respect to its meaning as the way the word looks on a page. Just as we all accept that a word spelled 'chair' could *in some language* mean basically anything, a word with exactly 'chair''s place in the linguistic structure could mean basically anything as well. If this is right, then just as

knowing how to spell words entails no understanding of those words' meanings, knowing how words fit into the linguistic structure of a language entails no understanding of those words' meanings. Therefore, learning a lot about how 'chair' ought to be used with other terms, as SuperMary learns, is insufficient for any understanding of the meaning of 'chair'. To achieve such understanding, one must connect 'chair' to a perception of a chair or to language that has already been tethered to perceptual content.

Here is a potential objection: Let's say we include mathematical language as part of "English". You read geometry textbooks to Mary and train her to answer questions and produce proofs of her own in the same way as you train her to respond to English queries. Then, you teach her to associate geometric formulae for cubes and spheres with 'cube' and 'sphere'. You may think that the arbitrariness that applies to standard English doesn't apply here. Perhaps the syntax alone guarantees the meaning of the symbols. If Mary is smart enough, can she thereby understand 'cube'? For this objection to succeed in threatening PF as a thesis about language that describes contingent states of the world, Mary's knowledge must potentially extend beyond abstract mathematical relations to a possible, physical cube. It is doubtful that Mary could somehow come to interpret the relevant mathematical formulae as descriptive of objects in physical space. However, if she *could* achieve this leap, presumably this will occur as she somehow comes to model the geometrical formulae in her head, such that she can imagine cubes under the descriptor 'cube'. This possibility does not threaten PF, but only offers a new way to achieve Ostensive Competency[32]. Without a

[32] If *Full Permissibility* from Chapter 1 — or even a much weaker alternative — is right, this possibility relies on Mary's learning the language of fellow humans who happen to perceive cubes as she does.

geometrical model of this kind, it is unclear what it could mean to say that Mary comes to think about cubes as possible, contingent physical objects in space, beyond roles in a mathematical system.

Finally, it is worth considering what SuperMary *does* understand. When people ask 'Do LLMs understand what they are saying?', the negative response often on offer says that LLMs, or the computers that run them, are just fancy autocomplete machines that manipulate symbols but understand nothing. Many can't shake the sense that, given an LLM's prowess with language, there *must* be some understanding present. Yet, while denying that LLMs understand anything about the *meanings* of our words, we might insist that they *understand* a lot about our language. For instance, they understand much about this peculiar artifact, the word 'chair', and how it relates to other words. While they cannot directly represent chairs, they *can* directly represent the word 'chair' and thus can learn a great deal about its properties.[33] Applying this interpretation to SuperMary and SuperBari: their understandings of 'chair' are sensibly indiscernible because the understood properties of 'chair' are the same in English and Alienese.

4. The Mentalese Dictionary

The previous section argued that one cannot learn the meanings of words merely by learning how those unknown words relate to one another. Yet, we can learn words from

[33] Given my focus on the relation between language and perception, I am concerned here with LLMs that deal only in language. If you accept *Full Permissibility* from Chapter 1, augmenting with image data doesn't change much. However, AI-powered robots that interact with the world (of the sort imagined by the "Robot Reply" to Searle's Chinese Room argument) are beyond the scope of my claims about LLMs.

those we already understand. If we restrict ourselves to languages that must be learned, the threat of regress requires that some words are not learned from other words. Thus, I've argued that some words must be learned from perception. However, in 1975, Fodor proposed the existence of an innate, unlearned "Language of Thought" ("Mentalese"), that does not itself require learning, that can facilitate the learning of natural language:

> My view is that you can't learn a language unless you already *know* one. It isn't that you can't learn a language unless you've already *learned* one. The latter claim leads to infinite regress, but the former doesn't . . . [T]he language of thought is known (e.g., is the medium for the computations underlying cognitive processes) but not learned. That is, it is innate. (65)[34]

While Fodor, like Wittgenstein, does not deny that Ostensive Learning is a standard part of language learning, Fodor's view entails the logical possibility that Mary might come to understand English entirely via the Dictionary Method by correctly translating new English terms into terms into "Mentalese" without developing any Ostensive Competency.

Here is Fodor's basic argument for Mentalese:

P1) To learn the meanings of words of their first natural language, children must form and confirm hypotheses about the words' meanings. (This is the only serious cognitive science explanation for the learning of natural languages.)

P2) To form a hypothesis about the meaning of word w before knowing the meaning of w, one must be able to represent the hypothesis without using w.

[34] Part of the omitted portion of this quote, preceding the final claim, includes the hedge that it "seems to [him] entirely plausible" that the view is right. As the book evolves, he continues on to defend the claim that the stated view *is* right.

P3) Representing a hypothesis about the meaning of w requires a language-like system of representation.

C) In order to learn their first natural language, children must be equipped with an unlearned — "innate" — language ("Mentalese").

I will not reject or endorse this conclusion, although I will argue that Mentalese cannot replace perception as the foundation for language learning. To appreciate Fodor's conclusion, the critical concept that must be understood is "innateness" or the quality of being "unlearned". At one extreme, we have the absurd thesis that mind-wandering fetuses may ponder snow, carburetors, and umbrellas (all which have their own atomic names in Mentalese, according to Fodor). On the opposite extreme, we have the truism that fetuses are such that, under suitable conditions, they will eventually be able to think about snow, carburetors, and umbrellas. Neither of these can be Fodor's claim about innateness. Fodor assures us babies only start out with finitely many concepts: "Minds like ours start out with an innate inventory of concepts, of which there are more than none but not more than finitely many." (131) Given that there is presumably an infinite variety of possible alternatives to our atomic concepts — imagine concepts for possible non-horses a bit different than horses, non-carburetor car parts a bit different than carburetors, etc. — such that we could have learned languages that coined terms for these alternative concepts, we must have the ability to acquire new concepts. Fodor grants that we can somehow acquire new concepts through experience, although he insists that this acquisition cannot be learning:

> Well, to acquire a concept is at least to know what it's the concept *of*; that is, what's required of things that the concept applies to. So, maybe learning the concept GREEN is coming to believe that GREEN applies to (all and only) green things; it's surely plausible that coming to believe that is at least a *necessary* condition for acquiring GREEN. Notice, however, that (assuming RTM) a token of the concept GREEN is a constituent of the belief that the concept GREEN applies to all and only *green* things. A fortiori, nobody who lacked the concept GREEN could believe this; nobody who lacked the concept GREEN could so much as *contemplate* believing this. A fortiori, on pain of circularity, coming to believe this *can't* be the process by which GREEN is acquired. (2008, 137)

Instead, some subintentional process is posited by which experiences of greenness or doorknobs trigger the development of corresponding concepts. So what does it mean to "have a concept"? Fodor adopts a referentialist semantics, such that the meaning of DOORKNOB is determined by the set of things that count as doorknobs and knowing the concept of DOORKNOB is knowing "what's required of things that the concept applies to" (2008, 137). But what does it take to know "what's required of" things that the concept applies to?

From reading and discussing Fodor, I learned the term 'carburetor', which I am told names a car part found in cars made before the late 1990s. To preserve this example, I have refrained from learning more about what exactly a carburetor does, or, in Fodor's locution, "what is required" of a thing to be truly called 'carburetor'. Even before being told that carburetors are car parts, I knew that 'carburetor' applies to all and only the carburetors. Perhaps this knowledge is problematically circular. Perhaps I can't have knowledge that *uses* the CARBURETOR concept until I've acquired the concept. Yet, I also know that 'carburetor' refers to all and only the things that are truly called 'carburetors' in English. This would be a circular way to *define* 'carburetor' in English, yet there is no paradox

involved in my knowing this true, if trivial, information about the term 'carburetor'. After all, presumably, I also have a mentalese concept of the *term* 'carburetor', and the relation between carburetors and 'carburetor' is merely contingent.

Given that the Fodorian concept acquisition need not be a rational process, there need not be an all-encompassing story about how it occurs. Conceivable candidates for mind-world interactions that cause concept acquisition are said to include "sensory experience, motor feedback, diet, instruction, first-language acquisition, being hit on the head by a brick, contracting senile dementia, arriving at puberty, moving to California, learning physics, learning Sanskrit, and so forth indefinitely."(2008, 131-2) Identifying the processes that do in fact cause concept acquisition is said to be a "wide-open empirical issue". (2008, 132) While you could acquire DOORKNOB from interactions with teapots, although this is not the standard process. Fodor grants that some story is required to explain how concept acquisition typically works:

> [S]urely acquiring DOORKNOB is a process that typically involves interacting with some of them; indeed, it typically involves interacting with *good examples* of doorknobs, where this means something like *stereotypic* examples of doorknobs. A viable nativism must have a story to tell about why that is so. (2008, 148)

To think about the conditions for acquiring CARBURETOR, consider this simple story:

> *Amy's Year of the Carburetor:* After encountering a new term, 'carburetor', in *The Language of Thought*, Amy asks Bob what a carburetor is. Amy, born when carburetors were already becoming obsolete, is generally quite ignorant of the workings of cars. She has never seen or touched a carburetor. Bob, an old car enthusiast, decides to take a few minutes each morning over the course of the

> subsequent year to enlighten Amy to the wonders of carburetors. On day one, Amy learns that 'carburetor' is an English noun. One day two, she learns that carburetors are material objects. On day three, she learns that carburetors were invented in the last 300 years. . . On day thirty, she learns that carburetors are car parts. . . On day forty, she sees a simple drawing of a carburetor. . . On day fifty, she sees a carburetor in the distance. Over the next month, she sees the carburetor a bit more clearly each day. Over the following month, she gets small daily lessons in the working of the carburetor. By day 365, she is competent in recognizing carburetors and as knowledgeable in the essentials of carburetors as any mechanic.

On day one, there are three possibilities regarding Amy's relationship to CARBURETOR, to be considered in turn: (1) Amy lacks CARBURETOR, (2) Amy has CARBURETOR but doesn't know that 'carburetor' means CARBURETOR, (3) Amy has CARBURETOR and knows that 'carburetor' means CARBURETOR.

Presume that Amy lacks CARBURETOR on day one. We are told that one typically gets a concept of a thing from good examples of that thing, and we can stipulate that Amy has never seen or touched a carburetor. Furthermore, Fodor is clear that CARBURETOR is "primitive", lacking a definition. The "definition" that Amy knows — x is a carburetor iff it is truthfully called a 'carburetor' in English — must therefore be insufficient, despite not using the term 'carburetor' and thus lacking any straightforward circularity. On day 365, Amy must have CARBURETOR, as she presumably understands the meaning of 'carburetor' just about as well as anyone, and, per Fodor, having CARBURETOR is necessary for learning 'carburetor'. Amy thus acquires CARBURETOR at some point during the year. We might

suppose that this acquisition is gradual: as Amy is gradually acquainted with carburetors, she gradually acquires CARBURETOR. Almost all properties are vague; perhaps 'having CARBURETOR' is just one more vague property. However, we need some story of this transition. On day one, Amy seems to think to herself, *huh, I wonder what a carburetor is . . . maybe a carburetor is some kind of plant or maybe a carburetor is some fashion accessory from the 70s. . .* According to Fodor, Amy's thoughts are in Mentalese. If Amy lacks CARBURETOR, what concept plays the role of *carburetor* in her thoughts? Maybe she uses a complex concept like ENGLISH REFERENT OF 'CARBURETOR'. Then, at some point during the year, she must replace REFERENT OF 'CARBURETOR' with CARBURETOR.

If we interpret Fodor's claims about Mentalese literally, then, for any given thought on any given day, there must be a fact of the matter about which concept was employed. It is one thing to say that the meaning of a term vaguely shifted over time; it is quite another to say that one term vaguely switched to another, especially when there are no candidates for intermediate terms. So, perhaps the switch happens at some point during the year. Is there anything more that can be said about this switch, or what would make it more or less likely for the switch to occur? We are told that the acquisition of concepts generally occurs via contact with stereotypical instances of that thing, although this is not a necessary condition. For instance, even though there are allegedly no actual triangles, you are more likely to get TRIANGLE from interactions with triangle-ish stuff. Presumably, we can sometimes get concepts of a given thing simply from learning about that thing via language, even without perceiving the thing or anything that perceptually resembles the thing in some special way. For instance, even without seeing diagrams, we could presumably learn enough about the liver to acquire LIVER; plausibly, most people acquire LIVER before ever

seeing a visual representation of the liver, which plays little to no role in the layperson's understanding of the liver. Prinz's (2011) objections to Fodor's account highlight the absence of a compelling explanation of how and why one should acquire concepts as they do.

For now, let's just grant that on some, hard-to-predict day in her *Year of the Carburetor*, Amy acquires CARBURETOR. Maybe on day 55, Amy sees the carburetor well enough to visually individuate key parts, causing her CARBURETOR acquisition. First thing in the morning on day 55, Amy seemed to think, *I wonder whether every car requires a carburetor.* After her CARBURETOR acquisition, Amy once again seems to think, *I wonder whether every car requires a carburetor.* According to the view we are considering, the latter thought uses CARBURETOR, while the former does not. Presumably, in the former case, she is using a concept like REFERENT OF 'CARBURETOR' instead of CARBURETOR. Now we can ask: why does Amy gain from the acquisition of CARBURETOR? What has changed? Will Amy's first personal experience of having this thought ostensibly about the ubiquity of carburetors differ in any predictable way after she acquires CARBURETOR? Does Amy immediately know how to translate 'carburetor', or REFERENT OF 'CARBURETOR', into CARBURETOR upon acquiring CARBURETOR?

Alternatively, suppose Amy *does* acquire CARBURETOR on day one, triggered by her encounter with 'carburetor'. She now has this new concept, CARBURETOR. She also has newly encountered 'carburetor', a string of English letters arbitrarily related to carburetors, and presumably has 'CARBURETOR'. Her possession of this CARBURETOR concept is allegedly critical to her coming to understand the English word for 'carburetor'.

Does she immediately know that 'carburetor' means CARBURETOR? Does she need to learn this via evidence? What sort of evidence could support such a translation? According to Prinz' 2011 review of *LOT 2*, the physical manifestations of Fodorian Mentalese concepts, just like the strings composing English words, should be understood as arbitrarily related to their referents. If Prinz is right, then Amy's brain state while tokening CARBURETOR is just a label that the human brain somehow uses to represent carburetors but might just have well come to represent anything. However, it cannot be the case that Amy could have used CARBURETOR to represent anything. We have already heard that it is a possible, empirically investigable hypothesis that someone might randomly acquire TEAPOT instead of DOORKNOB via doorknob interactions. Thus, it cannot be the case that whichever Mentalese label you acquire in conjunction with doorknob stimuli instantiates DOORKNOB. So when Amy acquires a new concept during a carburetor-ish experience — from first reading 'carburetor', from hearing carburetor information, from seeing carburetor diagrams, from perceiving carburetors — it is a live, if unlikely, possibility that she acquires ZAFU instead. (A zafu is a kind of meditation cushion.) If this happens, is it also possible that she will erroneously translate 'carburetor' to ZAFU and falsely come to believe that a ZAFU is a car part and that she sees a ZAFU during her carburetor lessons? Similarly, if Suzie randomly acquires TEAPOT from the doorknob, will she erroneously come to think that TEAPOTS are parts of doors? Would such errors ever lead to errors in Amy's or Suzie's use of English? This possibility of impactless inversion highlights the superfluousness of concepts on this view, as described by Prinz:

> Concepts are arbitrary symbols that can be used for nothing other than representing categories. They cannot be used to draw inferences, to plan actions, or to categorize.

> All of those functions are handled by the contents of our mental files. But this picture is very odd. It renders concepts needlessly anemic. (2005, 935-6)

We then return to the question discussed in Section 3. What makes it the case that a state of Amy's brain, or a "symbol" of Mentalese, comes to *represent* carburetors or anything else? On day one, Amy sees 'carburetor' in *LOT 2*. From context, Amy ascertains 'carburetor' to be an English noun. She immediately acquires a new Mentalese symbol, say '#'. '#' is arbitrarily related to carburetors. Yet, somehow '#' comes to instantiate CARBURETOR by referring to carburetors. How does this happen? At this point, the only connection between Amy's '#' and carburetors is 'carburetor'. One possibility is that upon seeing 'carburetor', Amy uses '#' for the very first time to think, *I wonder what a carburetor is.* This very first time Amy thinks about a carburetor — which is also the very first time Amy uses '*#*' — Amy (or Amy's brain) christens '*#*' as her term for carburetors. What determines that '*#*', in its first use by Amy, refers to carburetors rather than anything else? For instance, surely, the following story is at least coherent: Amy is looking at a magazine and encounters 'zafu' for the first time. A few minutes later, Amy picks up *LOT 2* and reads a few paragraphs, ending on 'carburetor' (also novel to Amy). Then, for whatever reason, distracted Amy thinks to herself, *I wonder what a zafu is.* Fodor endorses such a possibility by claiming that one could by happenstance acquire TEACUP from 'carburetor'. Yet, according to Prinz's interpretation, Amy's thought includes a novel Mentalese symbol acquired to refer to zafus/carburetors. That is, just after encountering both 'carburetor' and 'zafu' (and many other things) for the first time, Amy has the Mentalese thought T: I WONDER WHAT A _____ IS, with the blank occupied by a novel symbol '#'. One story says that '#' instantiates CARBURETOR in T. Another story says that '#' instantiates ZAFU in T.

What determines which is correct? Perhaps, upon "coining" '#', Amy somehow also stores some predicates in conjunction with '#', such as 'is called a 'carburetor' (in English)'. In this story, a belief featuring 'CARBURETOR' is critical to the acquisition of CARBURETOR. So how does Amy acquire 'CARBURETOR'?

Perhaps she has already acquired concepts of the letters and of English words, which can then be combined into 'CARBURETOR'. This solution seems suspicious for a few reasons. First, as she thinks about carburetors later that night, it is probable that she won't have the spelling of 'carburetor' perfectly committed to memory. Second, in the case that 'carburetor' is highly polysemous, it seems determinate that the meaning Amy is pondering about is the one relevant to the instance of 'carburetor' that caused her pondering. Perhaps, then, when Amy perceives 'carburetor', she perceives carburetor *as a novel word*, which her brain coins with '#', tagged in her brain to her perception of that 'carburetor' token. On this picture, it's not possible to get CARBURETOR randomly — e.g. from being hit on the head with a brick as Fodor suggests. The referential power of # is thus derivative of the referential power of 'carburetor'. Amy's brain is just using a symbol to represent *the referent of that word*. If this is right, then Amy cannot learn 'carburetor' by learning to translate it to CARBURETOR.

On this picture, it would be more accurate to say that each of us thinks in our own fully unique "Language of Thought", rather than that we all think in "*the*" Language of Thought. While Amy's brain uses '#' to refer to carburetors, John, who acquired CARBURETOR as a toddler watching his Dad fix cars, will use a different arbitrary mental symbol, say '$', for carburetors, and Mei, who acquired CARBURETOR from hearing the

word for carburetor spoken allowed in Chinese will use yet another arbitrary mental symbol, say '@', for carburetors. To say that S has CARBURETOR is then just to say that S has some symbol or another in their personal language of thought that refers to carburetors. There is nothing about the nature of the symbol itself that causes its representation of carburetors. Amy's carburetor symbol gains its representational power from the word that already represents carburetor. Similarly, if John acquires CARBURETOR before learning 'carburetor', some other representation must be employed to allow John's new arbitrary symbol to become CARBURETOR by representing carburetor. Causation is not enough: if a carburetor hits John on the head while sleeping, and John's brain enters a new state, or coins a new, arbitrary symbol of Mentalese, this new symbol does not thereby come to be *about* carburetors.

Fodor (1987, 1990) has suggested that a Mentalese symbol will instantiate CARBURETOR if there is a counterfactually robust causal law linking carburetors to the symbol. Firstly, we can again ask, when does Amy get into such a state? Early in her *Year of the Carburetor*, carburetors themselves do not seem to cause her to token any concept. After all, if she were confronted with a carburetor, she wouldn't recognize it as such. If 'carburetor' *does* reliably cause her to token the relevant concept, why doesn't this account entail that she has acquired 'CARBURETOR'? If she doesn't acquire CARBURETOR until late in the year, we are again left to question why we need to postulate such a concept. Finally, it is difficult to square a causation dependent account with the fact that I often interact with things without thinking about them or tokening their concepts (i.e. doorknobs, floors,

stairs, etc.) while I can have perfectly good, private thoughts about things divorced from any interactions with them.[35]

Alternatively, if this sort of causation does allow you to acquire CARBURETOR, I argue that a perceptual representation of the carburetor necessarily plays a mediating role. Let's stipulate that you get hit in the head with a brick, and thereby, randomly acquire CARBURETOR, as Fodor says is possible. According to this causal law account, your acquisition is explained by the fact that you somehow have found yourself in the situation in which carburetors reliably *cause* a certain brain state. Now, we can ask: by what mechanism is this causation carried out? Fodor treats perception as a typical, but not necessary part of the story.[36] Yet, some physical mechanism must be responsible for the presence of carburetors to reliably bring about a certain brain state. I will argue that, according to the most natural account of perception, it is impossible for such a process to occur without a perceptual representation of a carburetor or something that represents a carburetor. For carburetors to reliably cause a physical effect on your body, there must be a state in which your body registers the existence of the carburetor — or the existence of something which indicates the existence of the carburetor — that permits you to think about carburetors via arbitrary symbols. I claim that this state, in which your body is

[35] Beck (2018) convincingly distinguishes between perception and cognition on the of stimulus-dependence: If I perceptually represent doorknobs when none are present, I am malfunctioning. But I can cognize about doorknobs all I like regardless of their presence or absence. A causal dependence account of concepts is difficult to square with the Stimulus-independence of cognition.
[36] Notably, Fodor (1996) entertains a view in which perception plays a more critical role in concept in concept acquisition, which, while still anti-Empiricist, could be more easily reconciled with FP.

affected by the world, thereby registering information about what the world is like, will be appropriately characterized as a perceptual state.

To judge this claim, it is helpful to have in mind some generic conception of perception, beyond a list of the perceptual capacities that humans happen to have. Rogers' *Perception: A Very Short Introduction* describes perception as "the processes that allow us to extract information from the patterns of energy that impinge on our sense organs" (1). So what is a "sense organ"? Consider the scenarios in which a subject extracts information about the world from impingements upon their body. Will any such scenario fail to constitute perception on account of the impingement occurring to a part that does not qualify as a "sense organ"? If so, we would need a conception of sense organs independent of their role in perception, and no such conception seems forthcoming. If we happened upon a creature who reliably extracts information about the world from the impingements upon some unusual body part, we might thereby discover a new sense organ. If this is correct, we can simplify our definition of perception to require only that the relevant impingements occur upon the potential perceiver.[37] Accepting this notion of perception means that any robust causal mechanism tying a bit of the world to a mental state must involve perceptual representation. For there to be a law-like relation between a bit of the world and the state of my brain, the world being in that state must affect my body, causing me to enter the relevant brain state and thereby — on the Fodorian view under discussion

[37] Some characterize various "self-directed" human processes as perceptual — i.e. proprioception, balance, pain, emotional states — which seem to inform of the state of ourselves as opposed to the state of the external world and do not involve any obvious impingements upon sensory organs. However, reluctance to categorize such processes as perceptual is due to their potential objects rather than because the impingement occurs upon a part of a subject that is not properly "sensory".

— represent some state of the world. This process thus necessarily features the world impinging on my body so as to allow me to represent some aspect of the world. Any concept that I acquire through this process will therefore owe its representational potency to its connection to this perceptual state. And any language learned via translation into such a concept will thereby be properly associated with a perceptual state, preserving PF.

5. Conclusion

I have defended the thesis that anyone, human or otherwise, who understands the meaning of language must appropriately associate some of that language with perceptual representations. That is, for at least some bits of language, one must associate that language with perceptual representations of what that language describes.

Section 3 rejected the possibility of learning the meaning of a word merely by learning how that word interacts with other, still-to-be-learned words. No matter how much pure linguistic data one masters, the relationship between symbols and meaning remains arbitrary. Some qualitatively identical language learner on some alien planet could train on qualitatively identical data, and yet, the meanings of the symbols in their context might radically differ from the meanings of the symbols in your context. Linguistic data alone thus cannot allow you to come understand one meaning at the expense of another.

Section 4 considered the possibility of learning the meaning of a word by translating that word into a term in Mentalese, a postulated mental "language" composed of arbitrary states that come to instantiate concepts. I argued that this scenario can only occur if the arbitrary brain state instantiating the Mentalese concept is properly associated with

perception, or with another concept that is, directly or indirectly associated with perception. Whether or not any sort of mental "language" plays a role in the learning of natural languages, perceptual representations are indispensable to the process.

I conclude that we cannot understand the meaning of an arbitrary symbol, linguistic or otherwise, merely by understanding how that symbol relates to other arbitrary symbols. Perception is required because perception offers us our only nonarbitrary representational connection to the world. When I see a red cube, I come to know that the cube looks *this* way to me. While the cube might looks radically different ways to other types of perceivers — as I've argued in earlier chapters — it is *not* arbitrary how the cube looks *to me*. At least some of the content of my perception will thus be shared by any of my qualitative duplicates. It is this type of content, which I do not need any further learning to grasp, that can serve as the foundation for all of my representational endeavors.

References

Chalmers, David J., 1996, The Conscious Mind: In Search of a Fundamental Theory, Oxford: Oxford University Press.

Davidson, Donald, 1987, "Knowing One's Own Mind", *Proceedings and Addresses of the American Philosophical Association*, 60(3): 441–458.

Fodor, Jerry A., 1975, *The Language of Thought*, New York: Thomas Y. Crowell.

——, 1987, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: The MIT Press.

——, (ed.), 80, *A Theory of Content and Other Essays*, Cambridge, MA: The MIT Press.

——,1996, *Concepts: Where Cognitive Science Went Wrong*, New York: Oxford University Press.

——, 2008, *LOT 2: The Language of Thought Revisited*, Oxford: Oxford University Press.

Ford, Jason., 2011 ,'Helen Keller was never in a Chinese Room', *Minds and Machines*, 21, 57-72.

Harman, Gilbert. 1990. "The Intrinsic Quality of Experience." *Philosophical Perspectives. A Supplement to Nous* 4: 31–52.

Jackson, Frank, 1982, "Epiphenomenal Qualia," *Philosophical Quarterly*, 32: 127–136.

Keller, Helen, 1990, *The Organization of Learning*, Cambridge, MA: MIT Press.

Merriam-Webster. (n.d.). Celebrate. In Merriam-Webster.com dictionary.

Nagel, Thomas, 1974, "What is it like to be a Bat?" *Philosophical Review*, 83: 435–456.

Peacocke, Christopher, 1984 [1997], "Colour Concepts and Colour Experience", *Synthese*, 58(3): 365–381. Reprinted in Byrne & Hilbert 1997a: 51–66 .

Prinz, Jesse, 2002, *Furnishing the Mind: Concepts and Their Perceptual Basis*, Cambridge, MA: MIT Press.

——, 2005, “The Return of Concept Empiricism.” In H. Cohen & C. Leferbvre (eds.), *Categorization and Cognitive Science*. Elsevier.

——, 2011, “Has Mentalese Earned Its Keep? On Jerry Fodor’s LOT 2”, *Mind*, 120(478): 485–501.

Putnam, Hilary, 1975, “The Meaning of ‘Meaning’,” in Keith Gunderson (ed.), Language, Mind and Knowledge (Minnesota Studies in the Philosophy of Science, Volumes VII), Minneapolis: University of Minnesota Press, 1975; reprinted in H. Putnam, Mind, Language and Reality (Philosophical Papers, Volume 2), Cambridge: Cambridge University Press, 1975.

Quine, W. V. O., 1960, *Word and Object*, Cambridge, MA: The M.I.T. Press.

Rapaport, W., 1984, ‘How Helen Keller Used Syntactic Semantics to Escape from a Chinese Room’, *Minds and Machines*, 16(4): 381–436.

Rogers, Brian, 2017, *Perception: A Very Short Introduction*, Oxford, England: Oxford University Press.

Russell, Bertrand. 1948. *Human Knowledge: Its Scope and Limits*. London: George Allen and Unwin.

Searle, John R. 1980. “Minds, Brains, and Programs.” *The Behavioral and Brain Sciences* 3 (3): 417–24.

Whiteley, C. H., 1956, “Meaning and Ostensive Definition,” *Mind*, 65: 332–335.

Wittgenstein, Ludwig, 1953, *Philosophical Investigations*, New York: Macmillan.