

9-2016

# Seeing and Perceptual Content

Ben S. Phillips

*The Graduate Center, City University of New York*

[How does access to this work benefit you? Let us know!](#)

Follow this and additional works at: [https://academicworks.cuny.edu/gc\\_etds](https://academicworks.cuny.edu/gc_etds)

 Part of the [Philosophy of Mind Commons](#), and the [Philosophy of Science Commons](#)

---

## Recommended Citation

Phillips, Ben S., "Seeing and Perceptual Content" (2016). *CUNY Academic Works*.  
[https://academicworks.cuny.edu/gc\\_etds/1405](https://academicworks.cuny.edu/gc_etds/1405)

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact [deposit@gc.cuny.edu](mailto:deposit@gc.cuny.edu).

# SEEING AND PERCEPTUAL CONTENT

Ben Phillips

A dissertation submitted to the Graduate Faculty in Philosophy  
in partial fulfillment of the requirements for the degree of Doctor  
of Philosophy, The City University of New York

2016

© 2016

BEN STUART PHILLIPS

All Rights Reserved

SEEING AND PERCEPTUAL CONTENT

Ben Phillips

This manuscript has been read and accepted for the Graduate Faculty in Philosophy in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Michael Devitt

---

Date

---

Chair of Examining Committee

Iakovos Vasiliou

---

Date

---

Executive Officer

Peter Godfrey-Smith

Gary Ostertag

Jesse Prinz

David Rosenthal

---

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

# ABSTRACT

## SEEING AND PERCEPTUAL CONTENT

Ben Phillips

Advisor: Jesse Prinz

There are two widely held assumptions about perception: *ascriber-independence* (the view that the facts regarding what a subject perceives, as well as what her perceptual states represent, are independent of the interests of those attributing the relevant states to her), and *determinacy* (the view that perceptual content is relatively determinate). I challenge both of these assumptions, and develop a new approach to perceptual content, with implications for theories of mental content more broadly. In chapter one, I address the question of whether, in addition to low-level features, vision represents ordinary objects. I argue that there is just no fact of the matter. In chapter two, I defend a contextualist account of object-seeing: one that illuminates the inscrutability thesis defended in chapter 1. Finally, in chapter three, I address the question of whether the contents of vision are object-dependent. I argue that it is simply indeterminate whether the particulars we perceive enter into the contents of our perceptual states. I then address various worries about the indeterminacy thesis, arguing that we should embrace the view that there are multiple, equally acceptable, ways to assign contents to our perceptual experiences.

# Preface

During the 1980s and 1990s, psychosemantics was in full swing. Big hitters such as Dretske, Millikan, and Fodor all attempted to naturalize semantic content by appealing to naturalistically respectable conditions. However, as things stand, there is a sense—at least in many circles—that the whole enterprise was somewhat of a failure. Of course, this is probably due, in part, to sociological factors: philosophy is certainly not immune to fads. Nonetheless, I think there is also a philosophical reason that psychosemantics lost momentum. In brief, I think there is a widespread suspicion that standard approaches to naturalizing mental content are beset by a systemic inability to overcome various indeterminacies. For instance, there is Quine’s inscrutability thesis; Kripkenstein’s rule-following considerations; Fodor’s disjunction problem; Devitt and Sterelney’s qua problem; and the various indeterminacies that purportedly plague teleosemantic theories of content.

In light of these persistent indeterminacies, some philosophers have called for a radically different approach. For instance, proponents of the so-called phenomenal intentionality program have suggested that we ground mental content in phenomenology. More strongly, proponents of the phenomenology-first approach often claim that the roots of content determinacy are grounded in phenomenology. I’m dubious of these sorts of claims for a number of reasons. More importantly, though, I think that pessimism about naturalistic theories of content is unwarranted. But that is not because I think that the indeterminacy worries mentioned above can all be resolved by adding bells and

whistles to pre-existing theories. Rather, it is because I do not regard many of them as serious worries to begin with. Philosophers of mind (and language) have—in my estimation—grossly underestimated the degree to which the content of a mental state can be indeterminate, whilst still playing the explanatory roles that we would want it to.

The issues that I pursue in this dissertation can thus be seen as forming part of a much larger project that aims to show why we should be willing to live with (a surprising degree of) indeterminacy. Vision is a good place to start, for it is certainly tempting to think that (barring unusual cases), our visual experiences are crisp and determinate. When you look at a ripe banana, it is easy to be seduced into thinking that your experience determinately represents a yellow, banana-shaped, thing. How could it be otherwise?! It is this kind of temptation that I aim to resist in what follows.

# Acknowledgments

In forming and polishing the views that appear below, I have been helped by a number of gracious people. I would very much like to thank Jacob Berger, Tony Dardis, Ryan DeChant, Michael Devitt, Nemira Gasiunas, Peter Godfrey-Smith, Zoe Jenkin, Uriah Kriegel, Laura Larocca, Oliver Larocca, Myrto Mylopoulus, David Neely, Gary Ostertag, Adam Pautz, Jesse Prinz, Jake Quilty-Dunn, Jesse Rappaport, Adriana Renero, David Rosenthal, Jonathan Schaffer, Susanna Schellenberg, Henry Shevlin, Elmar Geir Unnsteinsson, Miguel Varela, and Iakovos Vasiliou. Special thanks to my advisor, Jesse Prinz, for the invaluable feedback and guidance he has provided. I am also very grateful to David Rosenthal, Michael Devitt, Peter Godfrey-Smith, and Gary Ostertag, who all served on my prospectus committee and gave me very useful feedback on early versions of the material below. I am especially grateful to David, who gave me very helpful input on some nascent ideas surrounding the inscrutability of reference. I am also especially grateful to Jacob Berger, Uriah Kriegel, Gary Ostertag, and Elmar Geir Unnsteinsson, who all provided extensive feedback on draft versions of the chapters that appear below—our discussions have been instrumental in shaping my overall project and its methodology.

BP

*New York*

*1 October 2015*



# Contents

<i>List of Figures</i>	xii
<b>Introduction</b>	<b>1</b>
<b>1 Object-seeing and the inscrutability of visual reference</b>	<b>7</b>
1.1 Introduction	7
1.2 Evidence for object-representations in vision	10
1.2.1 Amodal completion and visual search tasks	10
1.2.2 The automatic spread of attention	12
1.2.3 The neural basis of amodal completion	13
1.2.4 Volume completion	14
1.2.5 Multiple-object tracking (MOT)	16
1.3 The inscrutability of visual reference	18
1.3.1 Quinean inscrutability	19
1.3.2 Amodal completion, the spread of attention, and inscrutability	20
1.3.3 MOT and inscrutability	21
1.3.4 Re-identification and inscrutability	24
1.3.5 Solidity and inscrutability	26
1.4 Visual inscrutability and cognitive consumption	26

1.4.1	Fodor’s argument for the scrutability of reference	27
1.4.2	Applying Fodor’s strategy to thoughts about 3D objects	29
1.4.3	Fodor’s strategy and visual reference	29
1.4.4	The concepts-all-the-way-down argument	30
1.4.5	The argument from demonstrative thought	32
1.5	Inscrutability and vision-for-action	33
1.5.1	Object-manipulation and visual inscrutability	34
1.5.2	Biological constraints to the rescue?	39
1.6	Visual inscrutability and first-person access	41
1.7	Is object-seeing inscrutable?	43
1.7.1	Indirect seeing and contextualism	44
1.8	Conclusion	46
<b>2</b>	<b>Contextualism about object-seeing</b>	<b>47</b>
2.1	Introduction	47
2.2	Initial evidence for context-sensitivity	48
2.3	A contextualist account of the folk notion	49
2.3.1	A scalar analysis?	51
2.3.2	The heterogeneity of ascribers’ interests	52
2.3.3	Context-sensitivity and surface-seeing	55
2.3.4	More evidence for context-sensitivity	57
2.3.5	Subject-sensitive invariantism?	59
2.3.6	Looking ahead	63
2.4	Contextualism and the anchoring of <i>de re</i> thoughts	63

2.4.1	Siegel’s regimented notion of object-seeing	64
2.4.2	Contextualism and reference-fixing for <i>de re</i> thoughts	65
2.4.3	Contextualism and the differentiation constraint	67
2.5	Contextualism and vision science	72
2.5.1	Problems with the revisionist view	73
2.5.2	The ambiguity view	74
2.6	Arguments for the context-sensitivity of the scientific notion	75
2.6.1	Individuating biological traits and individuating states of seeing	75
2.6.2	MOT and explanatory interests	78
2.7	Conclusion	83
<b>3</b>	<b>Object-seeing and the generalism versus particularism debate</b>	<b>84</b>
3.1	Introduction	84
3.2	The basic varieties of generalism and particularism	86
3.2.1	Generalism	86
3.2.2	Particularism	88
3.2.3	Strong particularism	88
3.2.4	Weak particularism	89
3.2.5	Adjudicating the debate	90
3.3	Explaining phenomenology	91
3.3.1	The generalist’s argument from phenomenology	91
3.4	Intuition pumps and accuracy conditions	94
3.4.1	The particularist’s argument from intuitive accuracy conditions	94
3.4.2	Problems with the argument from intuitive accuracy conditions	94

3.4.3	Tye's argument from looks reports	96
3.4.3.1	Tye's reply	98
3.5	Demonstrative thoughts and knowledge of particulars	100
3.5.1	The argument from the one-many	101
3.5.2	The two-vehicle alternative	103
3.5.3	Contextualism (again)	105
3.5.4	The strong particularist's reply	106
3.5.5	The epistemological argument	108
3.6	There is no fact of the matter	110
3.6.1	Intuition pumps and salient content assignments	112
3.6.2	The irrelevance of parsimony	114
3.6.3	The multiple contents view	117
3.7	Context-sensitivity vs. indeterminacy	118
3.8	Conclusion	120
	<b>Bibliography</b>	<b>122</b>

# List of Figures

1. Amodal completion	10
2. Visual search task: version 1	11
3. Visual search task: version 2	11
4. The automatic spread of attention	13
5. Partially occluded bar	14
6. Volume completion	15
7. The multiple-object tracking paradigm	17
8. Intersection of square and triangle	27
9. Partially occluded face	82

# Introduction

What kinds of objects do we see? Suppose you are watching a plane that is about to disappear behind a thick black cloud. At present, the plane is in “full view”: that is to say, the cloud is not occluding any part of the plane just yet. What do you see? Do you see the plane itself? Do you only see its facing surface? Perhaps you see both of these things. If you *do* see both of these things then what about the other fusions of which they are arguably parts? For instance, do you see the object that is composed of the plane and Vladimir Putin’s left nostril?

Now consider what you see once the front half of the plane has been occluded by the cloud. Do you still see the plane itself? Do you only see its back half? Perhaps you only see the facing surface of its back half. What do you see when the only part of the plane that is not occluded by the cloud is the very tip of its tail?

In addressing these kinds of questions, philosophers have—unsurprisingly—offered an array of different answers. Both Broad (1925, 149) and Moore (1953, 34) argue that, strictly speaking, we only ever see the facing surfaces of objects. On their view, to say that one sees ordinary 3D objects is just loose talk.

Jackson (1977) disagrees, arguing that we do see ordinary 3D objects. This is because, according to him, we see them *in virtue of* seeing their “reasonably substantial” parts (e.g. their facing surfaces). More recently, Uriah Kriegel (2009) has defended a

similar view according to which one sees an ordinary 3D object just in case one sees a “highly integrated” part of it.

Other theorists, such as Pylyshyn (2003, 2007) and Burge (2009, 2010), agree that we see 3D objects; but that is not because we see them in virtue of seeing their parts. Rather, it is because the visual system *represents* them. According to Pylyshyn (2003, 2007) the visual system contains a “visual-indexing” mechanism, whose job it is to keep track of environmental particulars by maintaining referential contact with them over time. Importantly, these indexes refer to objects without encoding any of their properties: they are therefore akin to pure demonstratives (e.g. “this” and “that”). On this version of the representational view, seeing an object is tantamount to visually indexing it; however, as we shall see, Pylyshyn is not exactly clear on what kinds of particulars are referred to by visual indexes.

Burge (2009, 2010) denies that vision picks out environmental objects by deploying *pure* demonstratives, arguing that in order to visually perceive an object, one’s visual system must accurately encode some of its properties. More specifically, according to Burge, in order to see a 3D object, one’s visual system must, at the very least, accurately represent that object *as* a cohesive, bounded, and spatiotemporally continuous body.

Even though the approaches outlined above contain genuine insights, I think that there is an important ambiguity in the answers they give to our original question: namely, “What kinds of objects do we see?” The ambiguity stems from a failure to clearly distinguish between the following key questions:

- (1) What do our visual states represent?
- (2) What are the consequences, if any, for the notion of object-seeing that earns its explanatory keep in *vision science*?

- (3) What are the consequences, if any, for the *folk notion* of object-seeing?
- (4) What are the differences, if any, between the folk notion of object-seeing and the one that earns its keep in vision science?

In answering these questions, I challenge what I take to be two widely held assumptions about visual perception. According to the first, which objects one qualifies as seeing is an ascriber-independent affair: that is to say, the interests of ascribers have no role to play in determining which environmental particulars the subject qualifies as seeing. According to the received view, whether *S* counts as seeing an environmental particular, *O*, is solely determined by the head-world relations that obtain between *S*'s visual experience and *O* (e.g. causal relations). I argue that both the folk notion of object-seeing, and the one that earns its explanatory keep in vision science, are interest-relative.<sup>1</sup>

According to the second assumption, visual states have relatively determinate extensions. I argue that visual states are indeterminate in two important, and surprising, ways. First, I argue that there is no fact of the matter as to whether they represent objects, or, just their facing portions. More carefully, I argue that the contents of visual states are indeterminate as between the following candidates: ordinary 3D objects; the surfaces of ordinary 3D objects; and the *facing* surfaces of ordinary 3D objects. Second, I argue that there is no fact of the matter as to whether the objects we see enter into the

---

<sup>1</sup> The notion of object-seeing that I'm concerned with in this dissertation is to be clearly distinguished from other notions of seeing. These other notions include *fact-seeing*, *property-seeing*, and *phenomenological seeing*. Fact-seeing involves seeing *that* something is the case, e.g. seeing that Trump's hair is flailing in the wind. Property-seeing involves seeing properties of objects, e.g. seeing the shape and motion of trump's hair. Both fact-seeing and property-seeing are factive, whereas, phenomenological seeing is not. For example, even though Macbeth was hallucinating, he still qualified as "seeing" a dagger before him in the phenomenological sense of the term. See Dretske (1969, 1990) for a detailed discussion of these other notions of seeing.



contents of our visual states: in other words, it is indeterminate whether visual states have object-dependent contents.

Rejecting the assumptions described above is theoretically interesting in its own right: as we shall see, traditional debates in the philosophy of perception are premised on the claim that they are true. But rejecting these assumptions is also interesting because, as I will argue, there are important connections between interest-relativity and indeterminacy. It is my contention that one of the reasons that the indeterminacies I'm positing have gone unnoticed is that the explanatory interests of ascribers (including both vision scientists and the folk) often make a given content-assignment—among the group of equally acceptable ones—*salient*. What looks like a visual state with determinate content is, in fact, a state with various equally acceptable interpretations, one of which has been made salient in the ascriber's explanatory context. In fact, to my mind, this tendency to mistake a merely *salient* content-assignment for a uniquely correct one is endemic to the philosophy of mind (but that is a topic for a future project!).

The structure of the dissertation is as follows. In chapter 1, I address the question of what our visual states represent. In doing so, I start by reviewing empirical evidence for the thesis that vision houses object-representations—not just representations of low-level features, such as colors, 2D-shapes, texture, and motion. In particular, I review studies of amodal surface completion, volume completion, multiple-object tracking, and vision-guided action, arguing that the empirical data do not adjudicate between the various candidate-extensions mentioned above: namely, 3D objects; surfaces of 3D objects; and facing surfaces of 3D objects. Given that the (facing) surface of an object is an undetached part of it, my view bears a close resemblance to Quine's (1960, 1969) thesis that reference is inscrutable: a thesis that is almost always pitched at the level of language and thought. I argue that even if reference is scrutable at the level of language and thought, it does not follow that *visual* reference is scrutable as well.

Having argued for the inscrutability of visual reference, I then draw consequences for the nature of object-seeing. If visual reference is inscrutable, it seems to follow that there is no fact of the matter concerning which of the candidate objects (listed above) we see. I argue that even though this is a surprising thesis, we can explain the intuition that we determinately see ordinary objects, as well as their facing portions, by construing the folk as contextualists about object-seeing: as far as the folk notion goes, we determinately see ordinary objects and their facing portions, but only relative to an ascriber's context.

In chapter 2, I go on to provide a detailed defense of the view that the folk are contextualists about object-seeing. In particular, I argue that, according to the folk, whether one counts as seeing  $O$  over and above its facing portion varies according to the interests of those ascribing the state of seeing to  $S$ . In fact, I go ever further than this, arguing that whether  $S$  qualifies as seeing the facing portion of  $O$  is a context-sensitive affair as well. Having made my case for the context-sensitivity of the folk notion of object-seeing, I then address two challenges. According to the first challenge, only a context-*insensitive* notion of object-seeing is fit to play the explanatory roles that philosophers of perception have traditionally been concerned with. According to the second, in order to ascertain what object-seeing is really like, we must uncover the notion that earns its keep in vision science, and there is no guarantee that the notion we uncover will be a context-sensitive one. I argue that neither challenge is compelling: the folk notion of object-seeing and the one that earns its keep in vision science are both context-sensitive.

Finally, in chapter 3, I address the question of whether perceptual content is object-dependent: an issue I remain neutral on in chapters 1 and 2. According to generalists, the content of a perceptual state is completely independent of the facts concerning which objects, if any, the subject happens to perceive. Similarly, the generalist also

claims that whether the subject's state counts as accurate or not is completely independent of these same facts. Particularists disagree. According to them, the accuracy conditions of the subject's state can only be satisfied by objects she perceives. In fact, some particularists go even further than this and endorse the stronger claim that if the subject perceives an object then that object enters into the content of her experience. I argue that neither side of the generalism/particularism debate is right. What generalists and particularists are providing us with are equally acceptable ways of assigning contents (and accuracy conditions) to our perceptual experiences. It is simply indeterminate whether the objects we perceive are represented by our perceptual experiences.

# Chapter 1

## Object-seeing and the inscrutability of visual reference

### 1.1 Introduction

Imagine you are ten-pin bowling. Your gaze is centered on the middle pin as you lower your arm to release the ball. What exactly do you see? Do you see the pin itself? Do you merely see its facing surface? Do you see both? The question of whether we see ordinary objects or just their facing surfaces is an old one. Broad (1925) and Moore (1953) both endorse the view that, strictly speaking, we only ever see facing surfaces. Jackson (1977) defends the intuitive view that we see *both* ordinary objects and their facing surfaces: this is because, for Jackson, we see ordinary objects in virtue of seeing their facing surfaces. Clarke (1965), Dretske (1969), and Neta (2007) all give contextualist answers: relative to some contexts, you count as seeing the object itself, while relative to others, you only count as seeing its facing portion.

Related to the question of what we see is the question of what our visual states represent. In the last few decades, findings in vision science have been converging on the view that the visual system represents objects, not just an array of low-level features. For instance, in a series of studies, Pylyshyn (2001, 2003, 2006, 2007) has argued that our capacity to track multiple objects is best explained by positing representations of

them in early vision. Further evidence comes from the fact that amodal completion—the process whereby we represent objects as having occluded (and unseen) parts—seems to occur in early and mid-level vision, before object recognition takes place (Rensink & Enns 1998; Sugita 1999). The environmental particulars that are represented by states that participate in these visual processes are often referred to as “visual objects.”

These two questions—the question of which kinds of particulars we see, and the question of whether the visual system contains object-representations—are connected in an important way. For there are compelling reasons to think that seeing an object requires visually representing it in some manner or other (Dretske 1969; Burge 2009, 2010). On some versions of the representational view, seeing an object amounts to picking it out via a pure demonstrative (Pylyshyn 2003, 2007). On other versions of the view, it is necessary to accurately represent the object as having certain properties (Burge 2009, 2010). If we adopt a representational account of seeing then the evidence, mentioned above, concerning the kinds of entities that our visual states represent can be brought to bear on the old question of whether we see objects, facing surfaces, or both.<sup>2</sup>

In this chapter, I will focus mainly on the question of what our visual states represent, for doing so will enable us to shed light on the traditional question of whether we see objects or just their facing surfaces: the issue I focus on in chapter 2. In particular, I will argue that as far as the empirical evidence goes, there is just no fact of the matter as to whether our visual states represent objects, their facing surfaces, or both. I will do

---

<sup>2</sup> Of course, not everyone agrees that perceiving an object requires visually representing it in some manner or other. So-called *relationists* (e.g. Campbell 2002, Travis 2004, Brewer 2007) deny that visual states are representational in the first place; rather, they are, fundamentally, relations to environmental entities. On this view, to see an object is to stand in a non-representational relation to it. Schellenberg (2010, 2011) has argued—persuasively, I think—that one can accommodate the claim that perceptions are, fundamentally, relations to environmental objects, without denying that they are representational, for a perceptual representation might pick out an object via a demonstrative-like element. I take up the issue of whether visual states represent objects via demonstrative elements in chapter 3: in this chapter, I’ll remain neutral.

so by defending a variant of Quine's (1960, 1969) famous argument for the inscrutability of reference. Quine's thesis is almost always discussed as a thesis about the extensions of terms in language and thought. Famously, Quine argues that the term 'rabbit' is referentially indeterminate as between the property of *being a rabbit* and the property of *being an undetached-rabbit-part* (two necessarily co-instantiated properties). However, similar considerations can be used to argue that there is no fact of the matter as to whether we visually represent ordinary objects or just their facing surfaces: after all, the facing surfaces of objects are undetached parts of them.

If visual reference is inscrutable, and we accept a representational view of seeing, it would seem to follow that there is no fact of the matter as to whether we see ordinary objects or just their facing surfaces. But that is a surprising conclusion: it certainly seems as though we determinately see both ordinary objects and their facing surfaces. I allay this concern by offering two explanations of the folk intuition—explanations that are compatible with the inscrutability of visual reference.

The structure of the chapter is as follows. In sect. 2, I review empirical evidence for the claim that we visually represent objects (not just an array of low-level properties, such as colors). In sect. 3, I argue that the data do not adjudicate between the view that the states in question represent objects, and the alternative according to which they represent just their facing surfaces. In sect. 4, I address the worry that if thoughts have determinate extensions, perceptual states must have determinate extensions as well. I argue that there is no compelling reason to accept this inference. In sect. 5, I respond to the worry that we can only explain the fine-grained actions that we perform on objects if the visual states that mediate them determinately pick out ordinary 3D bodies. In sect. 6, I address first-person considerations, arguing that they do not provide us with any evidence against the visual inscrutability thesis. Finally, in sect. 7, I draw conse-

quences for the traditional debate concerning whether we see objects or just their facing surfaces, anticipating the detailed discussion of this issue in chapter 2.

## 1.2 Evidence for object-representations in vision

Over the past few decades, findings in vision science have been converging on the view that the visual system contains representations of cohesive, bounded, and spatiotemporally continuous particulars—not just representations of low-level features such as color, shape, texture, and motion. In what follows, I review what I take to be the most compelling pieces of evidence accrued by vision scientists.

### 1.2.1 Amodal completion and visual search tasks

Consider figure 1. In (a) you see a notched circle, whereas, in (b), you seem to see a complete circle, partially occluded by a square. This process of perceiving objects as having occluded parts is known as amodal completion.



**Fig. 1** Amodal completion

Rensink and Enns (1998) carried out an experiment in which they asked subjects to search for the notched circle among a field of distractors. For instance, try to find the notched circle in figures 2 and 3 below.



**Fig. 2** Visual search task: version 1



**Fig. 3** Visual search task: version 2

What Rensink and Enns found is that it is harder to locate the notched circle in 2 than it is in 3. They found that in 3, the task is easy, yielding a parallel search outcome: that is to say, there was little or no increase in search time as the number of distractors increased. In contrast, they found that 2 yields a serial search outcome: that is to say, there was a linear increase in difficulty as the number of distractors increased. Why is this the case?

The most straightforward explanation is that amodal completion is a pre-attentive



process that occurs in parallel vision.<sup>3</sup> When you examine figure 3, early visual processes parse the scene into an array of visual objects containing a notched circle: that notched circle is thus available for attention to glom on to. On the other hand, when you examine figure 2, early visual processes have parsed the scene into an array of visual objects that only contain complete circles (and squares): the missing part of the notched circle has been filled in. As far as the visual system is concerned, there is just no notched circle for attention to glom on to in figure 2.

For our purposes, the key point is that amodal completion appears to be a perceptual phenomenon. This is evidenced by the fact that it seems to occur pre-attentively, as a result of parallel processing: a type of processing that is widely seen as a marker of early perception.<sup>4</sup> Thus given that amodal completion involves the representation of *objects-with-partially-occluded-parts*, this amounts to evidence for the view that the visual system represents objects, not just an array of low-level features.

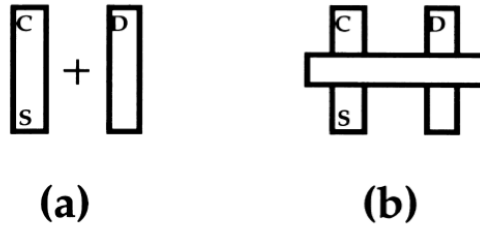
### 1.2.2 *The automatic spread of attention*

Another piece of evidence for the view that vision houses object-representations concerns the automatic spread of attention. Consider figure 4 below.

---

<sup>3</sup> The term “pre-attentive” must be used with caution. Single cell recordings in animals (Desimone & Duncan 1995) and functional imaging in humans (Kastner & Ungerleider 2000) have shown that many levels of processing (including processing within primary visual cortex) can be modulated by top-down selective attention. See Raftopoulos (2009, ch 1) for a very careful discussion of the relationship between attentional mechanisms and the representation of objects in early vision—Raftopoulos' approach is amenable to the inscrutability thesis that I defend below.

<sup>4</sup> For discussions of the widely held view that parallel processing is a key marker of early vision, see the following: Treisman (1985); Raftopoulos (2009); and Healey & Enns (2012).



**Fig. 4** The automatic spread of attention

In a series of trials, Egly, Driver & Rafal (1994) cued the subject to the end of one of the bars (labeled ‘C’ above). Immediately after being cued, the subject’s task was to detect a darkening at the end of one of the bars. Subjects were much faster at detecting the darkening if it occurred at the other end of the *same* bar (marked ‘S’), rather than the end of a different bar (marked ‘D’).

The fact that the distance from C to S is the same as the distance from C to D shows that there is a same-object advantage. What is happening is that the subject’s attention is spreading out from the cue to the boundaries of the object in question, which is further evidenced by the fact that it happens even if the bar that contains the cue is partially occluded (as in 4(b)). Again, the most straightforward explanation is that objects are represented by the visual system pre-attentively, before any of them are selectively attended to for further processing.

### *1.2.3 The neural basis of amodal completion*

Further evidence for the thesis that vision contains object-representations comes from investigations into the neural correlates of amodal completion. Numerous studies suggest that amodally completed figures are represented by bottom-up cell activations early on in visual processing.

For instance, using single cell recordings, Sugita (1999) found that orientation-

selective cells in the primary visual cortices of Japanese monkeys are sensitive to facts concerning occlusion. Subjects were presented with two vertical lines, separated by a grey patch (as in figure 5).



**Fig 5.** Partially occluded bar

When the patch was presented in such a fashion that it appeared to lie behind (or on the same plane as) the vertical line segments, the cell in question did not respond. However, when the patch was presented in such a way that it appeared to lie in front of the two line segments—so that it appeared to occlude a single bar—there was a significant response. Moreover, the response was similar to the one obtained when the cell was exposed to a single unoccluded bar. Sugita (1999, 271) concludes that “border completion is carried out in very early stages of visual processing.” Further studies utilizing both single-cell recordings and brain imaging all converge on the view that amodal completion is a visual process.<sup>5</sup>

#### 1.2.4 Volume completion

The studies I have described so far all involve 2D surface completion, but what about 3D volume completion? If the only compelling evidence for visual object-representations

---

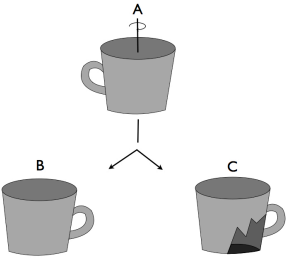
<sup>5</sup> For further studies utilizing single-cell recordings, see Bakin *et al.* (2000), as well as Bushnell *et al.* (2011). For a recent brain imaging study, see van Lier & Gerbino (2015).

concern representations of objects’ front surfaces, one might be tempted to side with Broad (1925) and Moore (1953) in drawing the conclusion that these are the only kinds of particulars that we ever see. However, there are compelling reasons for thinking that we visually represent the volumetric properties of ordinary 3D objects.

For instance, when you look at a bowling ball, its front surface occludes its rear surface. Nonetheless, from a phenomenological point of view, it is plausible that you visually experience the ball’s sphericity—a volumetric property that requires a form of amodal completion over and above 2D surface completion.

In contrast to the principles of 2D surface completion—which have been widely studied—the principles of volume completion are less well understood.<sup>6</sup> Nonetheless, apart from its phenomenological plausibility, the claim that volume completion is a *visual* phenomenon is well supported.

One reason for construing volume completion as a visual phenomenon concerns the fact that we complete the rear surfaces of self-occluding objects in a way that is imperious to our beliefs about them. For instance, consider figure 6 below.



**Fig. 6** Volume completion

---

<sup>6</sup> See Nakayama *et al.* (1995) for an influential account of surface completion. For an influential account of volume completion, see Tse (1999). Tse defends a unified account of volume and surface completion in terms of principles of “merging volumes.” On his account, surface completion is simply a degenerate case of volume completion (1999, 64).

Even if, upon looking at A, you know that the rear surface of the mug is the way C depicts it as being, you cannot help but see its surface as continuing on smoothly around the back to enclose a mug-shaped volume (as depicted in B). This insensitivity to countervailing evidence suggests that volume completion is not a matter of acquiring post-perceptual *beliefs* about the rear surfaces of objects; rather, it is a process that occurs within perception, before background beliefs are brought to bear.<sup>7</sup>

Another piece of evidence for the claim that volume completion is a visual phenomenon comes from a consideration of the action-guiding role of vision (which I discuss, at length, in sect. 1.5). The famous patient D.F. cannot recognize most 3D objects (a condition known as visual agnosia), and yet her ability to act on them remains intact to a significant degree: for example, she can pick up ordinary objects.<sup>8</sup> But successfully picking up ordinary objects requires representations of their volumetric properties: unless D.F. represents the 3D shape of a cup, say, how can she wrap her fingers around it in the requisite fashion? This suggests that volume completion occurs within vision, upstream of those processes that are responsible for object-recognition.<sup>9</sup>

### 1.2.5 Multiple-object tracking (MOT)

Perhaps the most compelling evidence for the existence of visual object-representations comes from studies of multiple-object tracking (MOT). In a typical MOT experiment, subjects are shown a screen containing up to 8 identical figures. Initially, some of these figures are flashed so as to mark them off as figures-to-be-tracked. The subject is then

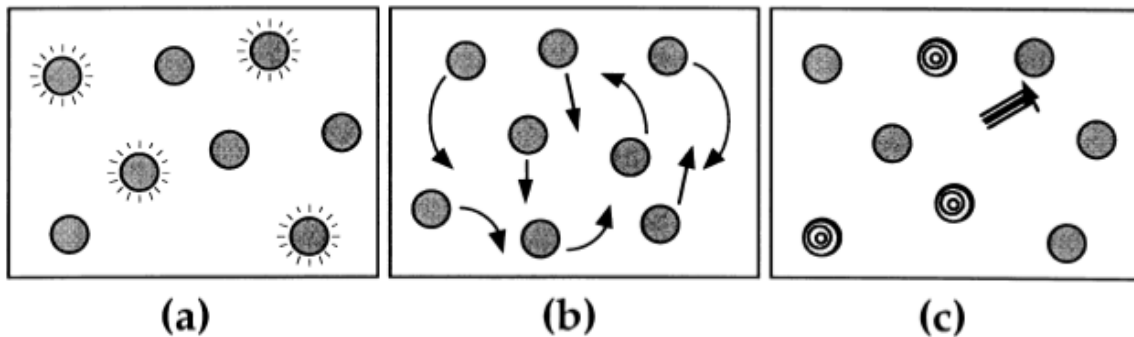
---

<sup>7</sup> See Nanay (2010) for an extended version of this argument.

<sup>8</sup> See James *et al.* (2003).

<sup>9</sup> See Tse (1999, 65) and Kubiluis *et al.* (2014) for similar arguments. See Marr & Nishihara (1978), as well as Biederman (1987), for influential accounts of the volumetric representations that serve as inputs to the process of object-recognition. Some theorists (e.g. Bayne 2009) claim that object recognition is a decidedly perceptual capacity: in what follows, I will remain neutral on this issue.

asked to track these targets as all 8 figures move along independent trajectories (see figure 7). At some point (5 to 10 seconds later) the motion stops and the subject is asked to indicate which figures were targets. Pylyshyn (2003, 223–32) has found that subjects are able to track 4–5 objects with an efficiency characteristic of parallel processing. However, beyond 5, the capacity falls off sharply, with a profile characteristic of serial processing.<sup>10</sup>



**Fig. 7** The multiple-object tracking paradigm

The fact that a normal subject’s capacity to track up to 5 objects has a profile characteristic of parallel processing suggests that the objects in question are being represented in early vision. Moreover, studies strongly suggest that subjects track objects independently of their color-, size-, shape-, and kind-properties. If subjects were tracking objects by visually representing these sorts of properties, we would expect performance to improve when targets and distractors are differentiated in the relevant respects: but this is not borne out by experiments. This suggests that the representations in question are being tokened early on in visual processing (certainly, before object recognition occurs).

In another study, Creem and Proffitt (2001) showed that if the semantic memory of normal subjects is tied up, their capacity to track objects is unaffected. For instance,

---

<sup>10</sup> See Howe *et al.* (2010) for a study that directly tests, and supports, the claim that MOT occurs via parallel processing.

subjects were asked to track a moving target on a screen while performing a semantic recall task. Performing this distractor task failed to impair their ability to track the target object(s), suggesting that tracking is a pre-cognitive capacity.

Equally important as the fact that MOT appears to be a visual process is the fact that if targets fail to behave in ways characteristic of physical bodies, tracking breaks down. For instance, even though objects can be successfully tracked amidst changes in surface color, size, shape, and kind, tracking breaks down if the target fails to maintain either cohesiveness, the integrity of its boundaries, or spatiotemporal continuity. For example, if a target starts to behave like a liquid, with portions breaking off in a pouring fashion, tracking breaks down.<sup>11</sup> Tracking also breaks down if the target fails to move along a spatiotemporally continuous trajectory, e.g. if it disappears and then suddenly reappears on the other side of the screen. Finally, experiments have shown that subjects cannot track easily defined *parts* of objects. Scholl, Pylyshyn and Feldman (2001) found that when a target is merged with a distractor via a connecting line, subjects are no longer able to track the target: instead, they track the composite object whose parts include target, distractor, and connecting line.

All of these constraints on successful tracking strongly suggest that the visual representations in play are keyed to cohesive, bounded, and spatiotemporally continuous particulars.

### **1.3 The inscrutability of visual reference**

The experiments outlined above all provide evidence for the view that vision parses the scene into an array of visual objects. But what exactly are the contents of the relevant states? Must we construe them as representations of 3D bodies? Why not construe them as representations of the facing surfaces of 3D bodies?

---

<sup>11</sup> See Yantis (1998); Scholl & Pylyshyn (1999); and vanMarle & Scholl (2003).

### 1.3.1 Quinean inscrutability

Given that the facing surfaces of objects are undetached parts of them, the question I'm posing here resembles Quine's (1960, 1969) famous question about the extension of the term 'rabbit.' Quine argues that there are several ways of assigning extensions to our terms that are compatible with all of the relevant data. Consequently, there is just no fact of the matter as to which assignment is correct: reference is inscrutable. For instance, there is no fact of the matter as to whether the term 'rabbit' refers to the set of rabbits, or, to the set of undetached-rabbit-parts: both candidates do equal justice to the relevant data.

More specifically, Quine's (1960, 1969) argument has the following basic structure. First, we settle the facts concerning the truth-values of sentences in the language. According to Quine, the data that determine these facts are speakers' dispositions to assent to, and dissent from, these sentences. We then go about assigning extensions to sub-sentential terms in such a fashion that these truth-values are preserved. This is where the inscrutability kicks in, for according to Quine, there are multiple (equally acceptable) ways to assign referents to sub-sentential terms such that the truth-values of the sentences they constitute are thereby preserved.

For instance, take the sentence, "There is a rabbit." This sentence is true in all and only those cases in which a rabbit is present. Thus we can assign the set of rabbits to the sub-sentential predicate 'is a rabbit.' However, given that *being a rabbit* and *being an undetached-rabbit-part* are necessarily co-instantiated, "There is a rabbit" is also true in all and only those cases in which an undetached-rabbit-part is present. We can therefore preserve the truth-value of "There is a rabbit" by assigning the set of undetached-rabbit-parts to the predicate 'is a rabbit.'

Before addressing the analogous thesis about *visual* reference, it is worth noting that many commentators reject Quine's argument on the grounds that it assumes an unac-



ceptable brand of linguistic behaviorism. For instance, Searle (1987) complains that Quine’s inscrutability thesis is premised on “extreme linguistic behaviorism,” the thesis that “the objective reality of meaning consists entirely of correlations between external stimuli and dispositions to verbal behavior” (1987, 126).

Now, my concern here is not with linguistic reference. Moreover, I will not contribute to the debate as to whether Quine’s original argument assumes an unpalatable form of linguistic behaviorism (I happen to think that it does not). Rather, I want to go about arguing that, as far as the relevant data are concerned, *visual* reference is indeterminate as between objects and their facing surfaces. In doing so, I will not assume any form of behaviorism. All I will assume is that (i) the relevant data are constituted by naturalistically respectable facts, and (ii) our epistemic access to those facts is naturalistically respectable as well.

With that in mind, I want to launch my argument for the inscrutability of visual reference by reconsidering the experimental findings reviewed above, for they constitute the primary source of evidence for the thesis that the visual system contains object-representations.

### *1.3.2 Amodal completion, the spread of attention, and inscrutability*

Reconsider Rensink and Enns’ (1998) experiment in which the subject searches for a notched circle among a field of distractors. Above, we said that the reason the subject finds it harder when presented with figure 2 is that her visual system is representing the target as a complete circle, not a notched one. But we can just as easily explain why she finds it harder by construing her visual system as representing the target as *the facing portion* of a partially occluded circle. In explaining why the subject finds it harder to locate the target in figure 2, the key point is that her visual system does not represent the target *as a notched circle* (i.e. as a pacman shape). Representing the target as a

(partially occluded) circle is one way for this to occur: representing it as the facing portion of a complete circle is another.

Similarly, reconsider the spread of attention. Above, we explained the same-object advantage exhibited in Egly, Driver, and Rafal's (1994) study by positing pre-attentive representations of objects over which the subject's attentional spotlight spreads. But we can just as easily explain the results by positing pre-attentive representations of the facing surfaces of objects: that is to say, we can construe subjects' attention as spreading out over the facing surfaces of objects, such as the bars in figure 4.

What goes for surface completion goes for volume completion too. For instance, when I look at the front surface of a sphere, there are two ways to assign an extension to my visual state. According to one, I represent the facing surface of the sphere *as such*: according to another, I represent the sphere *as such*. Both assignments enable us to explain my performances on those versions of the tasks discussed above, where the targets are 3D, as opposed to 2D.<sup>12</sup>

### 1.3.3 MOT and inscrutability

In assessing whether MOT is compatible with the inscrutability thesis that I'm seeking to defend, it is important to distinguish between representations of properties, and representations of the particulars that instantiate them (an issue I take up in much more detail in chapter 3).

Take Pylyshyn's account of MOT. According to him, the capacity to track multiple objects is enabled by a mechanism he calls "visual indexing" (2003, ch. 5). Importantly, these visual indexes refer to objects and maintain referential contact with them over

---

<sup>12</sup> It is important to realize that both 2D surfaces and 3D objects have facing portions. Thus even if surface completion occurs prior to volume completion in order of visual processing, both representations will exhibit the kind of indeterminacy that I'm positing.

time, but not in virtue of encoding their properties. Pylyshyn’s visual indexes are thus akin to pure demonstratives (e.g. “this” and “that”). Pylyshyn’s core motivation for construing visual indexes in this way stems from the finding—explained above—that tracking seems to proceed independently of those states that represent the color-, size-, shape-, and kind-properties of target objects.

Pylyshyn sometimes refers to visual indexes as “FINSTs,” which stands for “FINgers of INSTantiation.” He elaborates on this metaphor as follows:

... we initially viewed this mechanism in terms of the metaphor of keeping “fingers” on certain objects, so as to be able to refer to them, direct inquiries to them, or to move attention to them. If you imagine the cartoon character “Plastic Man” sticking long flexible fingers on a number of objects, and imagine that the fingers themselves cannot sense any properties of these objects directly but allows the individuated objects to be queried by attending to them, this captures the basic idea behind visual indexes. The term “instantiation” connotes the use of the indexes to bind mental particulars (or symbols) to objects, which then “instantiates” the variables. (2003, 206)

Does Pylyshyn’s account leave any room for inscrutability?

Suppose you successfully track the bowling ball as it rolls down the laneway. As the ball rotates, the part that constitutes its facing surface is constantly changing. How could it be that the visual index in question, say  $V$ , enables you to track the ball as it rolls down the laneway if what  $V$  picks out at each moment is changing? In order to explain how you successfully tracked the ball, don’t we need to construe  $V$  as maintaining referential contact with the same particular—namely, the ball—throughout?

Tempting as it is, I think this line of reasoning is mistaken. In assigning referents to visual indexes, the goal is to explain how a robust causal relation is maintained between

the visual system and objects in the world: even when, for example, those objects are rotating. And construing the referents of visual indexes as the facing surfaces of objects does nothing at all to undermine that explanatory goal. As long as the different facing surfaces that are successively picked out by a visual index are all facing surfaces *of the very same object*, the robust causal relation that we are after has been secured.

To borrow Pylyshyn's finger metaphor, suppose Plastic Man is keeping track of a spinning ball. His finger is always in contact with the ball, but the part that his finger is touching changes from moment to moment as the ball spins. It is precisely because Plastic Man touches different parts *of the same ball* as it spins that he is able to keep track of it. In the same way, we can explain how the visual system keeps track of a spinning object even if we posit an index that refers to different facing portions of that object at different points throughout the tracking episode: as long as they are all facing portions *of the very same object*, successful tracking has been achieved.

Even if one is not convinced by this reply, consider the fact that the entire surface of an ordinary 3D body is an undetached part of it. As the body rotates, the part that constitutes its facing surface changes, but the part that constitutes its entire surface does not. The kinds of worries raised above therefore do not even arise for this candidate referent.

What if we were to reject Pylyshyn's approach and construe visual indexes as referring to tracked objects by encoding their properties? For instance, according to Burge (2009, 2010), a visual index,  $V$ , refers to an environmental particular,  $O$ , just in case (i) there is an appropriate causal relation between  $O$  and  $V$ , and (ii)  $V$  accurately represents  $O$  as having certain features. Will any features do? Burge elaborates as follows:

Although shape, color, and kind are sometimes not tracked when objects are tracked, if the indexes pick out genuinely perceived objects they must (and do) carry mini-

mum coding of a perceivable type, however generic, that distinguishes figure from ground. (2009, 31)

Elsewhere, he elaborates as follows:

... spatial boundedness, spatial integrity, and continuity in motion are properties whose representation guides indexes for bodies. (2010, 456)

For instance, simplifying things a little, suppose  $V$  is of the form *THAT COHESIVE, BOUNDED, BODY*. In that case,  $V$  refers to  $O$  just in case  $O$  is a cohesive and bounded body, and it was also an appropriate cause of  $V$ .

If we adopt this alternative to Pylyshyn's view, we can easily explain MOT by construing the attributive elements of visual indexes as being referentially indeterminate as between ordinary bodies and their facing surfaces. For instance, according to the latter reference scheme, if I'm tracking a rotating sphere then  $V$  will be along the lines of *THAT FACING-PORITION-OF-COHESIVE-BOUNDED-BODY*—a complex demonstrative that picks out different facing portions of *the very same body* as it rotates.

#### *1.3.4 Re-identification and inscrutability*

One might worry that our capacity to re-identify objects during episodes of tracking can only be explained if we construe visual indexes as referring to objects themselves, not their facing surfaces. It is well known that humans, and various non-human animals, can track objects that disappear and then reappear from behind occluders.<sup>13</sup> For in-

---

<sup>13</sup> See Kellman & Spelke (1983); Spelke (1990); and Spelke, Kestenbaum, Simons, & Wein (1995).

stance, suppose that someone walks in front of me as I'm watching the bowling ball roll towards the pins. My view of the ball is momentarily blocked, but as soon as the offending party moves on, I'm able to re-identify the ball and continue tracking it. How could this be if the ball is rotating—the index in question thereby failing to pick out the very same object both before and after occlusion occurs?

There are various things to say in reply to this challenge. First, if the datum that requires explanation here is the capacity to track an object through temporary occlusion, I fail to see why the visual index in question must pick out the object itself as opposed to its (current) facing surface. As we have already seen, maintaining referential contact with whichever part of the object happens to be its facing surface suffices for tracking that object: the fact that tracking can survive temporary occlusion does nothing whatsoever to alter this point.

On the other hand, suppose the claim is that an ordinary subject has the capacity to visually represent an object as the same one that was experienced a moment ago. I'm not aware of any empirical evidence for the view that the visual system explicitly represents facts about the numerical identity of seen objects; however, even if we were to grant the claim that it does, this would not pose any problems for the inscrutability thesis. For instead of construing the subject's visual state as representing the given object, *O*, as being numerically identical to a previously experienced object, we can construe her experience as representing *O*'s (current) facing surface *as belonging to* the very same object whose facing surface was visible a moment ago.<sup>14</sup>

Once again, even if one is not convinced by this reply, notice that we still have the entire surface of the object as a candidate referent. Given that what constitutes the entire surface of the object does not change during rotation, there is no need to re-

---

<sup>14</sup> Quine (1969, 33) himself suggests this sort of re-interpretation of the identity predicate when accommodating speakers' answers to the question, "Is this *gavagai* the same as that?"

interpret the identity predicate in attempting to account for re-identification on this reference-scheme.

### *1.3.5 Solidity and inscrutability*

Finally, one might hold the view that we visually represent bodies *as solid*—that is to say, as entities that cannot penetrate (and pass through) one another (Baillargeon 1987; Carey 2009; Burge 2010, 465; Huntley-Fenner, Carey, & Salimando 2002). But the surfaces of objects are not solid bodies that resist penetration, so doesn't that rule them out as candidate extensions?

The reply to this worry should be familiar by now. As long as the facing surface of a solid body is represented *as such*—namely, *as the facing surface of a solid body*—that is all we need to accommodate the fact that normal subjects (as young as two months old) are surprised when the entities they are tracking pass through one another or behave like liquids and move in a pouring fashion.

Summing up: there is a preponderance of evidence for the claim that vision parses the scene into an array of “objects,” but the conception of objecthood in play does not adjudicate between 3D bodies, their surfaces, and their facing surfaces. As far as fixing visual reference goes, there is thus nothing to choose between these candidates.

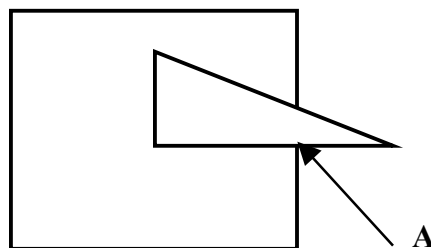
## **1.4 Visual inscrutability and cognitive consumption**

Up until this point, I have been focusing exclusively on *visual* states, but what about those states tokened downstream in cognition? Perhaps there are independent reasons for regarding them as possessing determinate extensions. If so, would that provide us with reasons for regarding the visual states that give rise to them as possessing determinate extensions too?

As I see it, there are three versions of this worry. According to the first, there are arguments for the view that linguistic expressions and thoughts have determinate extensions, which apply with equal force to the perceptual case. According to the second, visual object-representations are conceptual. Thus, if the concepts that are tokened in thought have determinate extensions, so must the concepts tokened upstream in vision. According to the third version of the worry, perceptual demonstrative thoughts inherit their referents from the perceptual states that cause them. Thus, if perceptual demonstrative thoughts have determinate referents, so must the perceptual states that give rise to them. Let's consider each worry in turn.

#### *1.4.1 Fodor's argument for the scrutability of reference*

In an influential reply to Quine's inscrutability challenge, Fodor (1994, 65) asks us to consider a figure of the following form:



**Fig. 8** Intersection of square and triangle

Notice that point *A* is an undetached part of both the triangle and the square. Thus, if we assign the set of undetached-triangle-parts to 'is a triangle,' the sentence "A is a triangle" comes out true. Similarly, if we assign the set of undetached-square-parts to 'is a square,' the sentence "A is a square" comes out true. The problem with this reference scheme, though, is that it fails to preserve the truth-values of other sentences in the



language. An ordinary English speaker will clearly reject the sentence, “A is both a square and a triangle,” and yet it is true on the non-standard interpretation. It follows that either ‘is a triangle’ does not have the set of undetached-triangle-parts as its extension or ‘is a square’ does not have the set of undetached-square-parts as its extension (or both).

As Fodor points out (1994, 74–7), this solution to Quine’s challenge assumes that purely informational theories of content are false, for it requires us to construe the inferential role of the predicate-conjunction symbol as (partly) fixing the extensions of our predicates. More specifically, we assumed above that because “A is both a square and a triangle” is false, either ‘is a triangle’ does not refer to the set of undetached-triangle-parts or ‘is a square’ does not refer to the set of undetached-square-parts. But there are non-standard reference-schemes on which “A is both a square and a triangle” comes out false, even though ‘is a triangle’ and ‘is a square’ are both assigned the non-standard extensions. According to one such scheme, ‘A’ is ambiguous, referring to the triangle in “A is both a square and a triangle”; to an undetached-square-part in “A is a square”; and to an undetached triangle-part in “A is a triangle.” This non-standard reference scheme preserves the fact that “A is both a triangle and a square” is false, even though ‘is a triangle’ and ‘is a square’ are both assigned the non-standard extensions. How can we rule out *this* reference-scheme?

Fodor (1994, 71) argues that we can do so by appealing to the fact that ordinary speakers will accept the inference from “A is a triangle” and “A is a square” to “A is both a triangle and a square.” If ordinary speakers accept this inference then it cannot be that “A” is ambiguous in the way envisaged above: if “A” were ambiguous in the way envisaged then speaker’s would reject the inference. It thus follows that ‘is a square’ and ‘is a triangle’ have determinate extensions, in part, because of the inferential role of the predicate-conjunction symbol—or so the argument goes—and that clearly constitutes a

departure from purely informational semantics, where the extensions of predicates are fixed solely in terms of the causal-informational relations in which they stand to properties in the world.<sup>15</sup>

#### 1.4.2 Applying Fodor's strategy to thoughts about 3D objects

In order to see how Fodor's strategy bears on *visual* reference, let's start by applying it to those thoughts that, on the standard interpretation, pick out ordinary 3D objects (not their facing surfaces). Take the mental predicate 'is a cube.' According to the standard interpretation, this predicate picks out the set of cubes; whereas, on the deviant interpretation, it picks out the set of facing-surfaces-of-cubes. Now, the predicate 'is a square' is clearly compatible with the deviant interpretation: something can be both square-shaped and the facing surface of a cube. However, it is clearly incompatible with the standard interpretation: nothing can be both square-shaped and a cube.

The problem for the Quinean is that if the cube in question is orientated in the right way, the sentence "That is both a cube and a square" will come out true on the interpretation according to which 'is a cube' has the set of facing-surfaces-of-cubes as its extension. We should therefore draw the conclusion that the predicate 'is a cube' does not have the set of facing-surfaces-of-cubes as its extension—or so the argument goes.

#### 1.4.3 Fodor's strategy and visual reference

In posing the question of whether Fodor's strategy can be used to undermine the thesis that *visual* reference is indeterminate, I will not take a stance on whether it works in the case of language and thought.<sup>16</sup> Rather, I will simply assume—for the sake of argu-

---

<sup>15</sup> See Fodor (1990) for a causal-informational theory of predicate reference.

<sup>16</sup> See Gates (1996) for a persuasive argument that Fodor's argument is unsound. See also Wilcock's (ms) persuasive critique, which develops an ingenious suggestion of David Rosenthal's.

ment—that Fodor’s argument is sound and that we therefore have determinate thoughts about objects and their facing surfaces. Does it follow that the representations tokened upstream in vision possess this same determinacy?

For instance, could we implement Fodor’s strategy by appealing to the fact that in looking at a cube, an ordinary subject will dissent (in thought) from the claim that *that* is both a cube and a square? The problem with this strategy is that it would only undermine the visual inscrutability thesis if we have already determined (i) which predicate-types are tokened in vision, and (ii) that they are the same ones that figure downstream in thought. Unless (i) and (ii) have been established, we do not have the leverage that we need in order to go from the fact that the subject dissents (in thought) from the sentence, “That is both a cube and a square,” to the conclusion that the visual states tokened upstream have determinate extensions.

In other words, we can only use Fodor’s strategy to argue that visual states exhibit the determinacy of thoughts if we reject the view according to which those object-representations in early vision that are implicated in MOT, and so on, are indeterminate as between objects and their facing surfaces; whereas, those states tokened downstream in cognition have determinate extensions. And the view that indeterminate visual representations give rise to determinate representations downstream in cognition (and language) is certainly coherent and plausible: in fact, it is a view that Fodor (2007) himself holds. More to the point, it is a view that takes work to argue against. Below I consider, and reject, two attempts to do so.

#### *1.4.4 The concepts-all-the-way-down argument*

According to the first argument, the contents of perception are conceptual in the sense

that they are apt to figure as the contents of cognitive states too.<sup>17</sup> If this view of perception is correct, what are the consequences for the visual inscrutability thesis? If we agree with Fodor in taking the concepts that constitute thoughts to have determinate extensions, and we also hold the view that visual states possess conceptual contents, there is a temptation to conclude that the concepts that figure in vision have determinate extensions as well. For instance, if we suppose that in looking at the blue sphere, the content of my visual state has the concepts *blueness* and *sphericality* as constituents, and we are also convinced by Fodorian-style arguments for the view that these concepts have determinate extensions, it seems to follow that perceptual states have determinate extensions as well.

The problem with this objection is that it just begs the question against the proponent of visual inscrutability. More specifically, it begs the question to insist that the concepts that are invoked in Fodorian-style arguments *are the very same ones* that figure in those stages of visual processing that parse the scene into cohesive, bounded, and spatiotemporally continuous particulars. According to the proponent of visual inscrutability, the empirical data do not adjudicate between the view that visual representations pick out objects, and the alternative according to which they pick out the facing surfaces of these objects. Moreover, according to the proponent of visual inscrutability, this holds true regardless of whether the visual representations in question are conceptual.<sup>18</sup> And if that is the case, it follows that those concepts—e.g. *sphericality*—that are appealed to in Fodorian-style arguments against Quine’s thesis are not the same ones that

---

<sup>17</sup> For a discussion of different ways to draw the conceptual/non-conceptual distinction, see Heck (2000, 2007) and Speaks (2005).

<sup>18</sup> Recall that studies of MOT and amodal completion strongly suggest that they are pre-attentive and parallel processes that occur in early vision. This remains the case, regardless of whether one rejects the additional claim that they are also *pre-conceptual* processes.

are deployed during those processes that mediate multiple-object tracking; amodal completion; and so on.<sup>19</sup>

#### 1.4.5 *The argument from demonstrative thought*

According to the second argument, perceptual demonstrative thoughts inherit their referents from the perceptual states that give rise to them.<sup>20</sup> Thus if we are already committed to the view that perceptual demonstrative thoughts determinately pick out ordinary objects (e.g. spheres) and determinately attribute properties to them (e.g. *being spherical*), it seems to follow that the perceptual states that give rise to them must have determinate extensions as well. More carefully, it seems to follow that perceptual states have singular elements that determinately pick out particulars, as well as attributive elements that determinately pick out properties: otherwise, how could the subject token a perceptual thought that determinately attributes a property to a unique object?

It is beyond the scope of this chapter to assess detailed views on how the referents of perceptual demonstratives thoughts are fixed. However, I want to suggest that we needn't do so in order to defend the visual inscrutability thesis against the worry at hand. For it is far from clear why a visual state with indeterminate referents could not give rise to a perceptual thought with determinate ones.

For instance, if something like Fodor's objection to the inscrutability thesis—as it applies to language and thought—is sound, that gives us all the resources we need to explain how a thought with determinate referents can arise from a visual state with indeterminate ones. To see why this is the case, recall that according to Fodor, no purely informational theory of content can secure referential determinacy: necessarily, terms

---

<sup>19</sup> Of course, another way to respond to the present objection would be to deny that the relevant visual states possess conceptual contents. That is certainly the view of theorists such as Pylyshyn (2003), Burge (2009, 2010), and even Fodor himself (2007).

<sup>20</sup> For example, see Martin (2002). I discuss this particular issue in more detail in chapter 3.

that carry information about rabbits also carry information about their undetached parts. What we need is a theory according to which the extensions of our terms are fixed, in part, by the inferences in which they participate. And, as we saw above, the crucial inferences for Fodor are those involving predicate conjunction. But if *that* is what fixes the determinacy of language and thought, we now have a content-determining story that is compatible with the thesis that indeterminate visual states give rise to determinate thoughts. According to the story I have in mind, whereas those visual states that mediate MOT, and so on, have indeterminate extensions, the determinacy of the *thoughts* that they typically cause are grounded in the inferential transitions that they participate in downstream of MOT and amodal completion—namely, inferences involving predicate conjunction and other logico-syntactic apparatus.<sup>21</sup>

### 1.5 Inscrutability and vision-for-action

One of the lessons from research on vision, action planning, and motor control in the last few decades is that an account of one is incomplete without an account of the others: vision evolved because it is in the service of action. In heeding this lesson, what are the ramifications for visual reference? Is the action-guiding role of vision compatible with the thesis that visual representations are indeterminate as between 3D objects and their facing surfaces? Perhaps the array of actions we perform on objects is too fine-grained for the visual inscrutability thesis to be tenable.

---

<sup>21</sup> Gareth Evans (1975) provides objections to Quine's inscrutability thesis that are very similar in spirit to Fodor's objection. In fact, Fodor (1994, 123) construes his own objection as a development of the insights found in Evans' article. Burge (2010, 216–23) and Wilcock (ms) provide what I take to be persuasive reasons for rejecting Evans' original arguments. But even if Evans' arguments rule out certain deviant reference-schemes, and even if they are sound when it comes to language and thought, nothing he says goes any way towards ruling out the thesis that the visual states we are concerned with (e.g. those that mediate MOT) are indeterminate as between 3D bodies and their facing surfaces.

### 1.5.1 Object-manipulation and visual inscrutability

Let's work with a simple example: you are thirsty, and so you pick up the cup of water that is sitting on the table. In order to explain how you perform this action, must we construe the *visual* states that mediate it as having determinate extensions? Do they determinately refer to the cup, its facing surface, or perhaps both? In order to answer these questions, let's trace out the key steps involved in performing the action.

First, you must select an appropriate target to perform the action on. In our example, this means that you must recognize the cup *as such*: a ventral stream process that is widely thought to implicate inferior temporal cortex.<sup>22</sup> Moreover, in order to recognize the cup *as such*, you must attend to it.

The second step involves the formulation of an appropriate action-plan: one that will achieve your goal of picking up the cup. Among other things, you must select which hand to use; an appropriate type of grip; and an appropriate type of posture. These are all choices that are made on the basis of processing within the ventral stream: bottom-up information is clearly insufficient by itself. The selection of initial grip and lift forces is informed by past experiences with objects of the relevant visual appearance.<sup>23</sup> For instance, my past experiences with cups, along with my recognition of the object *as a cup*, will aid me in selecting both an appropriate type of grip and an appropriate lift force, given my estimate of the cup's 3D shape and weight.

The fact that the detection of functional properties is involved in the planning stage makes it very likely that ventral stream processing is involved. Suppose the cup is, in fact, a coffee mug. Based on past experiences, as well as the fact that you recognize the cup *as a coffee mug*, you expect it to have a handle (one that is currently not visible).

---

<sup>22</sup> See Sato *et al.* (2013) for a discussion.

<sup>23</sup> Buckingham, Cant, & Goodale (2009).

You thus select an action that involves reaching around to the back of the mug where the handle is located.

Already, at the planning stage, we have the makings of a challenge to the visual in-scrutability thesis, for in attending to, recognizing, and representing key properties of the cup, aren't you tokening determinate whole-cup-representations? For instance, your representation of the cup's 3D shape is a key determinant of grip selection, and that is a feature of the cup, not its facing surface.

There are various things to say in response to this challenge. First, we need to distinguish between those object-representations that are decidedly visual, and those that are tokened downstream in cognition at the action-planning stage. The former are implicated in those segmentation and grouping processes that parse the scene into units that are fit for tracking; they also serve as inputs into mechanisms of object recognition, as well as other cognitive mechanisms. Thus, even if object-representations of the *cognitive* variety have determinate extensions, it does not follow that the object-representations tokened upstream in vision do as well—that was one of the morals we drew from §1.4.

What about those cases in which the subject successfully performs an action on an ordinary object without recognizing it? For instance, D.F. can pick up objects she fails to recognize: it is just that her impairment causes her to do so in non-functional ways (Carey, Harvey & Milner 1996). For example, she might pick up a hammer from the wrong end. Is it that, in such cases, her action is mediated solely by those visual object-representations that are tokened upstream of object-recognition? And if that is the case, don't these visual representations need to determinately represent the 3D-shape of the object she succeeds in picking up? I don't see why.

Suppose D.F. were to succeed in picking up the cup without recognizing it as such. As long as her visual system represents the cup's facing surface *as* the facing surface of a 3D body of such-and-such a 3D-shape, this will suffice for her to form an appropriate



action-plan. The key point is that the action-plan must be informed by a state that represents the presence of occluded object-parts, and this is something we get with a state that picks out the facing surface of a 3D body, and represents it *as such*.

Moreover, even if one is not convinced by this reply, notice that the problem raised above does not go any way towards undermining the thesis that D.F.'s visual state is indeterminate as between the cup and its *entire* surface, where the latter is a particular that clearly has occluded (and graspable) parts.

So the first and second stages of visuomotor action are compatible with the visual inscrutability thesis. Consider now the third stage, which involves a carefully controlled execution of those motor-programs generated as a result of the planning stage. In the case of grasping the mug, this will involve fine-grained adjustments of things like finger flexion, orientating of the wrist, and so on (components of reaching-and-grasping that occur relatively late in the performance).

Now, motor control is widely thought to be a largely dorsal stream process (implicating the parietal lobe); however, there is significant controversy regarding just how independent it is from object-recognition and planning, which are widely thought to be ventral stream processes. According to Milner and Goodale's influential model (1995, 2006, 2008), the dorsal processes responsible for control are largely independent of those ventral processes that underlie object-recognition and planning. In other words, once processing within the ventral stream results in the selection of both (i) a target object, and (ii) a type of action to be performed on that object, execution of the plan (involving specific motor programs and control mechanisms) is given over to the dorsal stream.

Critics have challenged this model, arguing that Milner and Goodale have underestimated the degree to which control is mediated by ventral processing.<sup>24</sup> For our purposes, the most important challenge comes from those theorists who argue for the fol-

---

<sup>24</sup> For instance, see Schenk & McIntosh (2010).

lowing combination of views: (i) control is mediated by conceptual representations, and (ii) those representations are thereby *visual*. If these two claims are correct then a familiar worry arises: if object-concepts have determinate extensions, and visual representations deploy object-concepts, then visual representations are constituted by object-concepts with determinate extensions.

For instance, consider Wayne Wu's (2008) argument for (i) and (ii). Wu starts by posing what he calls *the Many-Many Problem*: "there are typically many targets on which one can act, and for any target, there are many ways of acting on it." (2008, 1007) He elaborates as follows:

On the motor side, there are multiple levels of selection. First, there are many types of movement (reaching, throwing, etc.), only one of which can typically implement the intention. Secondly, within that type, one token movement (with specific trajectory, speed, force, etc.) must be selected and executed. Action requires that only relevant visual information guide this concrete movement. Identifying a one-one map at this level of precision, however, is not the result of practical deliberation. (2008, 1008)

What is it the result of then? According to Wu, it is the result of concept deployment:

Consider using a corkscrew ... During this process, multiple targets are presented and multiple manipulations of different parts of the corkscrew are needed. We are constantly visually registering new information. It is not that we deliberately form new intentions at each moment when additional information becomes relevant ... Rather we focus on the relevant information because we actively parse an object, tracking properties by shifting visual attention, all in the service of an intention to open the bottle with the corkscrew. The intention is, in a sense, the vehicle that brings the

relevant concepts to bear, concepts that are then independently exercised throughout action. In such cases, the exercise of concepts is a part of the visual process of selection. These visual processes are conceptual. (2008, 1020–21)

Why exactly does Wu claim that the concepts in question are deployed *within* vision? The basic thought is that—throughout action—concepts direct attention to those properties of the target object that are relevant to the action plan, including those stages that occur well after planning (i.e. those that concern control and the adjustment of specific motor programs) (2008, 1026).

I won't assess Wu's controversial claim that concepts are deployed within vision in the way he envisages. Moreover, I will even grant, for the sake of argument, that the concepts he appeals to have determinate extensions. For even if we grant both of these assumptions, it does not follow that the inscrutability thesis that I'm defending is false.

In order to see why, recall the evidence that we appealed to above in motivating the view that vision contains object-representations. In particular, recall the fact that the representations in question are tokened at those stages of visual processing that occur before capacities for object recognition are manifested. In fact, as we saw earlier, ordinary subjects track objects in a way that is insensitive to their color-, shape-, size- and kind-properties: all that is required is that they retain cohesiveness, boundedness, and spatiotemporal continuity. This is important, for it shows that early visual processing produces object-representations before the relevant objects are subsumed under high-level concepts, such as *CUP* and *HAMMER*. Thus even if Wu is right that objects are subsumed under high-level concepts *within vision*, it does not follow that the object-representations tokened upstream—the ones that serve as inputs to object-recognition—are constituted by the same high-level concepts. And if that is the case, it also does not

follow that the latter object-representations have the same (determinate) extensions as concepts such as *CUP* and *HAMMER*.<sup>25</sup>

### 1.5.2 *Biological constraints to the rescue?*

In addressing Quine's inscrutability thesis, Burge provides an argument from action that is potentially at odds with the claims made above. According to Burge, certain biological factors have a constitutive role to play in the determination of perceptual reference—in particular, those that concern basic actions, such as eating, mating, and so on (2010, 292). The key claim is that

... perceptual states are constitutively (partly) dependent for their representational content, not only on the environment's causally impinging on individuals, but on individuals fulfilling basic whole-animal functions. The constitutive ground for this latter dependency lies partly in the role that perception and perceptual kinds play in explaining realizations of individual biological function—centrally, individual activity. (2010, 371).

Applying this to Quine's inscrutability thesis, Burge argues as follows:

Bodies are more basic to biological explanations of most animals' pursuits than temporal stages, undetached spatial parts, or instances of universals (all as such). So bod-

---

<sup>25</sup> This will be important below (in sect. 1.7, as well as chapter 2) when we come to consider the issue of object-seeing, for the visual states that determine which objects a subject qualifies as seeing are plausibly those states that mediate MOT and amodal completion, not the states that are tokened downstream in the service of object recognition. In brief, the reason is that one can see an object without recognizing it, but one cannot see an object without differentiating it from its immediate surroundings (and that's what the processes of amodal completion and selection-for-tracking provide).

ies have prima facie priority in determining perceptual referents and contents. Most of the alternatives that Quine uses to suggest gratuitousness are ruled out by these sorts of considerations.<sup>26</sup> (2010, 215)

The biological constraints that Burge appeals to may well rule out some of Quine's candidates, but do they rule out *the facing surfaces* of 3D bodies? Do 3D bodies have priority when it comes to assigning visual referents? Given our discussion of action-guiding vision above, it is very hard to see why they should.

Even if we agree that grasping and eating an apple constitutes an action that fulfills basic whole-individual functions, that still leaves us with two equally good ways of assigning extensions to those visual states that inform the action-plan. According to one assignment, the visual representations in question pick out the *facing* surface of the apple and represent it as the facing surface of an apple-shaped body (with various occluded parts). According to another assignment, the representations in question pick out the entire surface of the apple, representing it as the surface of an apple-shaped body (with various occluded parts). According to yet another assignment, these representations pick out the apple itself. Nothing Burge says supports the view that the latter assignment puts us in a better position to explain how those states that are tokened further downstream in action-guiding vision result in successful acts of object-manipulation: acts that are crucial to individual biological function.<sup>27</sup>

---

<sup>26</sup> See Carey (2009, 99) for a similar argument.

<sup>27</sup> I take it that the same point applies to teleological theories of content (e.g. the kind of theory defended by Dretske 1986, 1988; and Millikan 1984, 2004). If the different reference-schemes described above each put us in an equally good position to explain successful acts of object-manipulation, then I fail to see how appeals to natural selection and biological function will adjudicate between them.

## 1.6 Visual inscrutability and first-person access

So far, I have been defending the thesis that visual reference is inscrutable by focusing on various capacities (both perceptual and cognitive). But what about first-person considerations? Might it be that the inscrutability of visual reference can be ruled out via evidence garnered from introspection?

Searle (1987) famously rejects Quine's inscrutability thesis by appealing to the first-person perspective:

... we know from our own case that we do mean by 'rabbit' something different from 'rabbit stage' or 'undetached rabbit part' ... In all discussions in the philosophy of language and the philosophy of mind, it is absolutely essential at some point to remind oneself of the first-person case. (1987, 126–7)

Might we rule out the thesis that *visual* reference is inscrutable by heeding Searle's admonishment?

Suppose the claim is that there is a phenomenological difference between visually experiencing an ordinary object and visually experiencing just its facing surface, where this tracks a difference in content. The problem, though, is that it is not at all clear what the phenomenological difference might amount to. How could there be any difference in what it's like to undergo a visual experience *as of a blue sphere*, and a visual experience *as of the facing surface of a blue sphere*? Moreover, even if there were a difference, how could we possibly tell? It is not as if we can cause a subject to undergo a visual experience as of a blue sphere, followed by an experience as of the facing surface of one, and then ask her whether there was a difference in phenomenal character.

In any case, even if we set aside appeals to phenomenology, I doubt that introspection can serve the desired evidentiary role. For suppose introspection is a reliable source

of evidence concerning the extensions of our visual states? In that case, assuming visual states have indeterminate extensions, what would we expect our introspective judgments to say about them?

Consider a simple example. *S* is gazing at the tip of an iceberg, the majority of which is submerged under water. Given our assumptions that (i) there are various, equally acceptable, ways of assigning extensions to her visual experience, and (ii) introspection is a reliable source of evidence concerning the extensions of our visual states, it seems reasonable to expect that *S*'s introspective judgment will be sensitive to whichever interpretation—among the group of equally acceptable ones—is rendered salient by her current interests. And that is arguably what we find.

If I ask *S* to characterize her visual experience, and she has just been wondering about the three-dimensional shape of the iceberg—something that cannot be deciphered based on the shape of its tip—she might utter something along the lines of “I can only see the tip: it looks white and pyramid-shaped.” On the other hand, if *S* is merely concerned with the location of the iceberg, it would be natural for her to report having a visual experience of the iceberg itself, not just the tip. For instance, she will say something along the lines of “I’m visually experiencing an iceberg as being over there.” Isn’t this variance among introspective reports what we would expect were there more than one acceptable way of assigning extensions to our visual experiences?

Of course, the friend of determinacy might deny that introspective judgments of the kind given above are reliable guides as to the contents of perception: perhaps they fail to reliably distinguish between the contents of perception, and the contents of immediate perceptual *belief*.<sup>28</sup> But if that is the case, why take them to be reliable guides as to

---

<sup>28</sup> See Pautz (2009) for some compelling arguments that we should not posit perceptual contents on the basis of the kinds of reports given above.

whether visual reference is inscrutable? That is to say, why heed Searle's advice and consult introspection in the first place?

### 1.7 Is object-seeing inscrutable?

If I'm right that visual reference is inscrutable, does it follow that it is indeterminate which kinds of particulars I see? More specifically, does it follow that there is no fact of the matter as to whether I see ordinary objects, or, just their facing surfaces?

One might try to avoid this consequence by denying that the facts concerning which objects *S* sees are fixed by the facts concerning which objects are referred to by her visual indexing system. Most obviously, one might appeal to the representations tokened downstream in the object-recognition system: if *those* representations have determinate extensions (and it is far from clear that they do!) then it might follow that *S* determinately sees ordinary objects. The problem is that that object recognition is just not a plausible level of processing to appeal to in fixing the facts concerning what *S* sees: in brief, doing so would belie the fact that one can see an object without recognizing it: an issue I explore in much greater depth in chapter 2.

Alternatively, one might try to avoid the consequence that object-seeing is inscrutable by denying that the facts concerning which objects one sees are determined by the facts concerning which particulars one's visual state *represents*, regardless of whether the states in question mediate MOT or processes downstream in objection recognition. For instance, perhaps we could adopt a purely causal theory of object-seeing: one that does not involve an appeal to what the subject's state represents.<sup>29</sup> The worry with this attempted dodge, though, is that the arguments for visual inscrutability offered above seem to apply with equal force to those accounts of visual reference—most notably, Pylyshyn's—that fall under the umbrella of causal theories of seeing. In any case, rather

---

<sup>29</sup> Grice (1961) is the *locus classicus* of causal theories of perception.



than going into that issue further, I want to conclude this chapter by briefly suggesting that the inscrutability of object-seeing is compatible with folk intuitions—the dodges suggested above are therefore unnecessary.

### 1.7.1 *Indirect seeing and contextualism*

As I have already noted, it is widely held that we see ordinary 3D objects indirectly *in virtue of* directly seeing suitable parts of them. For instance, I see the iceberg in virtue of seeing its tip; I see the house in virtue of seeing its façade; and so on. According to Jackson (1977, 17), the *in virtue of* relation is not a causal or inferential one: it is the analysis of one fact in terms of another (but not vice versa).<sup>30</sup>

If this kind of view is right then even if it is indeterminate whether my visual state picks out 3D object, *O*, and even if that implies that it is indeterminate whether I see *O* *directly*, it does not follow that it is indeterminate whether I see *O* *tout court*. This is because on each acceptable interpretation of my visual state, it either picks out *O*, or, it picks out part of *O* (e.g. its facing surface). Thus, on each interpretation, either I see *O* directly, or, I see it indirectly (in virtue of seeing one of its parts). On every acceptable interpretation, I thereby qualify as seeing *O*: what is varying with each interpretation—and is therefore indeterminate—is the manner in which I see it. There is thus a real sense in which I determinately see *O*, even though it is indeterminate whether my visual state represents it. And that seems to do justice to the folk intuition that we can determinately see ordinary objects.

Of course, one might counter that the notion of indirect seeing is problematic (e.g. perhaps it has no explanatory power). It is beyond the scope of this chapter to address

---

<sup>30</sup> What qualifies as a “suitable” part? Jackson (1977, 17) claims that it must be “reasonably substantial,” without elaborating on what he means by this. Kriegel (2009, 225) claims that the directly seen portion must be “highly integrated” into the relevant object. I discuss these issues—at length—in chapter 2.

that kind of concern; although, I am sympathetic to it.<sup>31</sup> But notice that if the folk are committed to a notion of indirect seeing, then that is all we need to explain the intuition that we determinately see ordinary objects. In any case, in chapter 2, I will argue that the folk notion of object-seeing is context-sensitive: that is to say, the folk are committed to the view that whether one sees *O* over and above its facing portion varies with the interests of ascribers. This paves the way for an alternative explanation of how the inscrutability of visual reference squares away with the folk notion of object-seeing.

In order to see why, suppose that the folk ascribe states of seeing in a context-sensitive manner. In that case, even if you think that the folk are mistaken—because the notion of seeing that does explanatory work in vision science is context-*insensitive*—that does not change the fact that the folk go around ascribing states of seeing in this way.<sup>32</sup> And if the folk do, in fact, deploy a context-sensitive notion of seeing, we have a ready explanation of why it might be that, intuitively, there are scenarios in which we qualify as determinately seeing ordinary objects; scenarios in which we qualify as determinately see their facing portions; and other scenarios in which we qualify as determinately seeing both. We can explain these intuitions because which option the folk go with in a given context depends on their interests as ascribers: relative to a given context, it may well be true that the subject sees an ordinary object (e.g. the plane), as well as its facing portion (e.g. its tail); relative to another context, it might be that the subject only sees the object's facing portion.

The key point is that if the folk deploy a context-sensitive notion of object-seeing then, according to them, we determinately qualify as seeing objects, and we also determinately qualify as seeing their facing surfaces: it is just that we only qualify as doing so

---

<sup>31</sup> See Phillips (2014) for worries about the explanatory impotence of the related notion of *indirect representation*.

<sup>32</sup> In fact, in chapter 2, I will argue that *both* the folk notion of seeing, and the one deployed in vision science, are context-sensitive.

relative to the relevant contexts. And this view is compatible with my claim that the visual *reference* is indeterminate as between objects and their facing surfaces. Which reference-scheme (among the group of equally acceptable ones) the folk go with will vary according to their interests as ascribers.

## 1.8 Conclusion

When it comes to thought and natural language, the thesis that we cannot determinately refer to ordinary 3D objects is counterintuitive. Moreover, the literature on why it might be false is extensive. On the contrary, the thesis that *visual* reference is indeterminate has received far less attention. However, it is not nearly as counterintuitive—if at all—and we should refrain from inferring the determinacy of perception from the determinacy of thought and language. In particular, we should resist the temptation to think that if reference is determinate at the level of language and thought, then perception is the source of that determinacy. It may well be that one of the key markers of the distinct between thought and perception is that, while the former determinately represents objects and their facing portions, the latter is subject to multiple, equally acceptable, interpretations.

Having argued that visual object-representations are indeterminate, in the next chapter I turn the focus more squarely onto object-seeing. In particular, I will elaborate on my contention that the folk deploy a context-sensitive notion of object-seeing, before arguing that the notion of object-seeing that earns its keep in the philosophy of perception and vision science is context-sensitive as well.

# Chapter 2

## Contextualism about object-seeing

### 2.1 Introduction

At the conclusion of chapter 1, I suggested that if we construe the folk notion of seeing as context-sensitive then we can account for the folk intuition that we determinately see objects and their facing surfaces—the reason being that if contextualism about the folk notion is true, that allows us to say that there are contexts relative to which there are determinate truths of the form “*S* sees *O*.”

In this chapter, I defend the view that the folk notion of object-seeing is, indeed, a context-sensitive one. Moreover, I argue that there are compelling reasons for thinking that the notions of object-seeing that earn their keep in the philosophy of perception and vision science are context-sensitive as well. In none of these cases can we talk about “the seeing relation”—at least, not in a way that isn’t tethered to the interests of an ascriber.

I proceed as follows. In sect. 2, I present the primary source of evidence for the context-sensitivity of the folk notion of object-seeing. Then, in sect. 3, I clarify the kind of context-sensitivity that this notion exhibits by contrasting it with the kind exhibited by those gradable terms (e.g. ‘tall’) that are widely seen as context-sensitive. Then, in subsequent sections, I address the suspicion that the context-sensitivity of seeing-ascriptions

is merely a feature of ordinary language. More specifically, in sect. 4, I reply to Siegel's (2006) contention that only a regimented, context-*invariant*, notion of object-seeing will figure in explanations of our capacity to form perceptual demonstrative thoughts. In sections 5 and 6, I argue that the notion of object-seeing that is deployed in vision science—most notably, in explanations of multiple-object tracking—is a context-sensitive one as well.

## **2.2 Initial evidence for context-sensitivity**

Consider the following cases. In each one, I describe two contexts: one relative to which it seems true to say that the subject sees the object in question, and another one relative to which it seems true to say that she does not.

### *Hide-and-seek*

Suppose I'm playing hide-and-seek with Oscar. I'm hiding behind the couch, but the tip of my nose is sticking out the side, thereby revealing my whereabouts to him. He utters the following: "I see you: your nose is showing!" On the other hand, suppose Richard is watching us play. He is eager to find out whether it looks, to Oscar, as though I have lost weight. He asks Oscar, "Do you think he looks slimmer?" to which I interject, "Well, let me get up from behind the couch first! Oscar can't see me right now: all he sees is my nose."

### *The close-up*

Suppose we are at a very crowded art gallery with Oscar. We have gone there to see a painting by a mutual friend, Laura. The gallery is so overcrowded that Oscar's eyes are an inch away from Laura's painting. You and I are wondering where Laura's painting is, and so we phone Oscar from another part of the gallery. He replies, "It's in the back

room on the left.” You question Oscar’s claim, to which he replies, “I know that’s where it is because I see it right now. It’s directly in front of my eyes!”

On the other hand, suppose Laura comes over and sees Oscar’s position, one inch away from the painting. She is very eager to get Oscar’s opinion about it, and so she utters the following: “Oh, he can’t actually see it right now. He’s way too close! All he sees is that small red patch in the corner.”

### *The cow show*

Suppose we are both in the audience at a contest in which our cow, Frida, is being evaluated. From the judge’s vantage point, every part of Frida is occluded except for her tail. You are very eager to see how Frida does in the contest and so you are annoyed that, from the judge’s vantage point, Frida is not in full view. You want proceedings to speed up and so you utter, “Ugh! The judge can’t see Frida right now: she can only see her tail!” On the other hand, suppose the judge is counting the number of cows that are present; there are 10 in total, including Frida. Having seen Frida’s tail, she turns to her fellow judge and utters, “I can see all 10 cows. Let’s begin the contest.”<sup>33</sup>

## **2.3 A contextualist account of the folk notion**

In each of the cases provided above, even though we have fixed the relations that hold between *S*’s visual experience and *O*, the truth-conditions of “*S* sees *O*” seem to be varying with the interests of ascribers. Thus if we take the cases given above at face value, what we get is the following basic view:

(CS) The truth-conditions of a sentence, *P*, containing the verb ‘sees’ vary according to the context in which it is uttered, independently of whether *P* is ambiguous or

---

<sup>33</sup> For some similar cases, see Dretske (1969, 27) and Neta (2007).

contains other context-sensitive terms. The contextual feature that is responsible for this variance in truth-conditions concerns the interests of the person who utters *P*. In particular, the interests of the utterer determine whether seeing a given part of *O* is sufficient for the subject to count as seeing *O* itself.<sup>34</sup>

There are at least three different ways to elaborate on (CS). According to the *surprise indexical view*, ‘sees’ is an indexical term, expressing different relations in different contexts of utterance (the utterer’s interest being the contextual parameter that determines the relation expressed). Alternatively, the *hidden indexical view* denies that ‘sees’ is indexical: rather, there is a covert variable in the logical form of any sentence containing the term ‘sees,’ where that variable takes on different values in different contexts, depending on the interests of ascribers. On the *unarticulated constituents view*, the literal/conventional meaning of a sentence, *S*, containing ‘sees’ falls short of determining a complete proposition: propositions conveyed by utterances of *S* thus have constituents that are not the values of any terms (overt or covert) in *S*. In what follows, I will remain neutral between these different ways of construing the context-sensitivity of

---

<sup>34</sup> Dretske (1969, 28) seems to have been the first to suggest a view along these specific lines, although he merely suggests it in passing without defending or clarifying the view in any detail. More recently, Neta (2007) has also defended a view along the lines of (CS); he too does not go beyond cases. Clarke (1965) defends a view that departs from (CS) in one important way. According to him, *sees* is a “unit concept.” A unit concept *F* only applies to *O* in a given context if *O* is the unit determined by that context, and it only applies to a certain amount of *O* (e.g. its facing surface) if that amount is the unit determined by the given context. It follows that if *F* applies to *O* in context *C*, then it does not apply to any part of *O*; whereas, if it applies to part of *O* then it does not apply to *O* itself. Thus, in the case of seeing, Clarke’s view is that if *O* has been singled out as “the relevant unit” then the subject counts as seeing *O*; whereas, if the facing surface of *O*, say *P*, has been singled out then the subject only counts as seeing *P* (not *O*). The problem with Clarke’s view is that it just does not fit the linguistic data. Seeing-ascriptions of the following sort are commonplace: “I see Bill *and* all of his toes”; “She saw the house *and* its beautiful façade.”

‘sees.’<sup>35</sup> Instead, I want to focus on the nature of those interests that the truth-conditions of seeing-ascriptions vary with.

### 2.3.1 *A scalar analysis?*

The fact that the ascriber’s interests are determining whether seeing a given part of *O* is enough for *O* itself to count as being seen invites a comparison with those gradable adjectives that are widely seen as context-sensitive: e.g. ‘tall,’ ‘flat,’ and ‘empty.’ Take the term ‘tall.’ It is widely held that this predicate expresses different extensions relative to different contexts of utterance, because the latter determine different minimum values that *O* must take on a scale of height in order to count as being tall. For instance, in the context of a discussion about basketball players, 5'10" does not count as tall, whereas, in the context of a discussion about jockeys, it does count as tall. This is often referred to as a *scalar analysis*, for we are explaining the context-sensitivity of ‘tall’ by positing a semantic link to a scale of height.<sup>36</sup>

Can we provide something like a scalar analysis for the context-sensitivity of ‘sees’? In answering this question, notice a key difference between the context-sensitivity of ‘sees’ and that of gradable adjectives such as ‘tall.’ The difference concerns the fact that ‘tall’ is semantically linked to a scale involving height, which goes hand-in-hand with constructions that specify and compare degrees of tallness; whereas, ‘sees’ does not seem to be semantically linked to a scale that measures *how much* the given object is seen, and as such, it does not permit constructions that specify and compare degrees of seeing. For example, on the only acceptable reading of “I see Venus a lot,” it conveys the proposition that I see Venus frequently: there is no reading on which it specifies the de-

---

<sup>35</sup> See Cappelen & Lepore (2005) for a critical discussion of these different accounts of context-sensitivity.

<sup>36</sup> See Kennedy (1999) for a detailed discussion of scalar analyses.



gree to which I see Venus.<sup>37</sup>

### 2.3.2 *The heterogeneity of ascribers' interests*

Could it be that 'sees' is semantically linked to a scale that measures *how much of* the object in question is seen, where context determines the minimum amount required for the object itself to count as being seen?

For instance, consider Jackson's (1977, 19) claim that a subject sees *O* just in case she sees a "reasonably substantial" part of it. What counts as a reasonably substantial part? Plausibly, that is a context-sensitive affair. If my interests as the ascriber concern your whereabouts then your nose counts as a reasonably substantial part; whereas, if my interests concern your posture then your nose will not count as a reasonably substantial part. Is the notion of object-seeing that emerges from Jackson's account therefore amenable to a scalar analysis?

To be sure, there are felicitous constructions that specify and compare how much *of* a given object is seen. For example, we do say things like the following: "I can only see half of the moon"; "I see more of the moon than I could last night"; and so on. Unfortunately, how much of a given object is seen does not seem to be the common factor that, relative to any given context, determines whether the subject counts as seeing the object itself. The interests at play are too heterogeneous for that to be the case.

As an illustration, recall the example in which I'm hiding from Oscar behind the couch, nose protruding. Relative to my interest in finding out whether I look slimmer, I utter the following seemingly true sentence: "Oscar can't see me: he only sees my nose." But what if I were to maneuver myself into a position such that part of my belly, and

---

<sup>37</sup> It is for this reason that Stanley's (2004, 2005) arguments against the view that 'knows' exhibits the same kind of context-sensitivity as gradable terms such as 'tall' do not threaten the contextualist account of 'sees' that I'm defending here.

that part alone, were now protruding? Moreover, suppose that enough of my belly were visible for a reliable judgment to be made about whether I have slimmed down. In this scenario it would seem true for me to say, “Oscar sees me.” But suppose I change positions once more, the result being that only the back of my head is visible. Moreover, let’s suppose that Oscar now sees more of me (by volume, weight, etc.) than he could when part of my belly was protruding. Does he count as seeing me now? Given my interest in having my slimness assessed, I seem to speak truly in uttering, “Oscar can’t see me: he only sees the back of my head.”

What this example shows is that context does not determine how *much* of *O* must be seen in order for *O* itself to count as being seen; rather, it determines which *kind* of part must be seen in order for *O* itself to count as being seen. Applying this to our example above, the claim is that relative to my context of utterance, *S* only counts as seeing me if my facing portion is the kind of object the seeing of which would enable *S* to determine whether I have lost weight.

The same kind of problem besets Kriegel’s (2009, 225) attempt to pin down conditions for when seeing part of *O* suffices to count as seeing *O* itself. According to Kriegel, if *S* sees *X*, and *X* is *highly integrated* into *Y*, then *S* counts as seeing *Y*. For instance, if I see the facing surface of an apple then I count as seeing the apple itself because the former is highly integrated into the latter. On the other hand, in seeing the apple I do not qualify as seeing the fusion of it and Kim Jong-un’s left shoe because the apple is not highly integrated into this fusion. Kriegel concedes that the notion of high integration is “vague and obscure”; however, he remarks that “something *like* the above analysis is probably right and more precisely expressible.” (2009, 227)

I agree that the notion of high integration is vague, but I also think it is likely context-sensitive. Moreover, I think it is probably context-sensitive in the same way as gradable terms, such as ‘tall’ and ‘flat.’ For instance, suppose I’m in a living room con-

taining a couch, various chairs, a coffee table, a lamp, and so on. Does the lamp count as being highly integrated into the arrangement of objects that is the living room? If my interest is in finding a comfortable place to sit then, presumably, I speak truly in denying that the lamp is highly integrated into the living room. On the other hand, suppose the living room I have wandered into is a movie set. Moreover, the set designer has painstakingly chosen where to place each object so as to realize her artistic vision for the relevant scene. Given *her* interests, it seems true for her to say that the lamp is highly integrated into the complex object that is the living room (at least, more integrated than it is relative to my interests).

If I'm right that the notion of high integration is context-sensitive, it follows from Kriegel's view that whether seeing part of *O* suffices for the subject to count as seeing *O* itself is a context-sensitive affair. In particular, it depends on the contextually determined standard for *being highly integrated into*, and whether the part in question meets that standard. For instance, if only the lamp is visible to you—perhaps you see it from afar through a window—then you count as seeing the room itself relative to the set designer's context of utterance. On the other hand, given my interest in finding a nice place to sit, you only count as seeing the lamp relative to my context of utterance.

Thus Kriegel's view arguably commits him to contextualism about object-seeing. However, the problem is that the interests that determine whether the subject counts as seeing a given object often have nothing to do with degrees of integration. For instance, Oscar's utterance of "I see you: your nose is showing" counts as true relative to the given context because his interest is in locating me. That is to say, all Oscar cares about is whether the part that he sees enables him to discern my whereabouts: whether that part is highly integrated into me is irrelevant for these purposes.

Thus the moral is as follows: the interests that the truth-conditions of ordinary seeing-ascriptions vary according to are a heterogeneous bunch, concerning matters such

as: the location of the object in question; which categories the object falls under; who the object is; whether the object's aesthetic qualities can be appreciated visually; and so on. Any theory that implicates only one of these interests will therefore fall short of accommodating the linguistic data.<sup>38</sup>

### 2.3.3 Context-sensitivity and surface-seeing

Up to this point, I have been arguing that whether seeing the facing portion of *O* suffices to count as seeing *O* itself is a context-sensitive affair. What I have not argued for is the thesis that whether one counts as seeing the facing surface of *O* varies with the ascriber's context. It is certainly tempting to think that if you count as seeing the facing surface of an ordinary object then this is a context-invariant fact about you. One way to express this tempting view is in terms of Jackson's (1977) distinction between *direct* and *indirect* seeing. Recall that according to Jackson, whenever one sees an opaque 3D object, one sees it indirectly in virtue of directly seeing its facing surface. Thus we can restate the view under consideration as follows: which objects we qualify as indirectly seeing is a context-*sensitive* affair; whereas, which objects we qualify as directly seeing is a context-*insensitive* affair. Tempting as it is, I think this view is mistaken.

First, notice that we sometimes count as seeing a three-dimensional object without

---

<sup>38</sup> Peter Godfrey-Smith (2005) has argued that we should construe folk psychology as a model—in brief, a model is an idealized structure whose resemblances to the target system allow us to represent that system in a way that generates predictions and explanations of its behavior. Godfrey-Smith's core claim is that when we engage in folk-psychological interpretation, we are not deploying a theory; rather, we are bringing a model to bear on the subject. Importantly, there are multiple construals of the model on offer (i.e. multiple ways to use the model as a representation of the subject): which one we go with will depend on our explanatory/predictive interests as folk psychologists. I take contextualism about the folk notion of object-seeing to be compatible with Godfrey-Smith's model-based account of folk psychology: given its amenability to different construals, it should come as no surprise that the folk psychologist's model engenders a context-sensitive notion of object-seeing. See Godfrey-Smith (2006) for a general discussion of model-based strategies in scientific theorizing.

seeing its facing surface. For instance, suppose I'm looking at Mars in the night sky. It appears to me as a point of red light, and as such, I cannot make out any of its surface details. For example, when asked if I can see any of its craters or mountains I reply, "I can't see its surface at all: it just looks like a point of red light to me." Importantly, in denying that I see the surface, I do not deny that I see Mars itself—I assert that I see it "as a point of red light." Thus it follows from cases like these that, as far as the folk notion goes, we sometimes see three-dimensional objects without seeing their facing surfaces.

Is there a context in which I count as seeing the facing surface of Mars, even though I cannot see it *as such*? Suppose the facts concerning whether I can make out any of its surface details are not at issue for *you*. For instance, suppose you are explaining why it is that Mars appears as a point of red light to me. You explain how light from the sun is reflected from its facing surface, travels through space, and then through the Earth's atmosphere to my retinas. In this context it seems true for you to say that I'm seeing the facing surface of Mars. After all, the light that passes through the atmosphere and impinges on my retinas was transmitted from Mars' facing surface!

Another example: suppose we are driving in the countryside, looking for a house with a very famous façade. The original house was built in the 1820s, but the façade is the only surviving part: the rest of the house having been replaced over the years. I gaze towards the house on the horizon, which appears to me as a very small and indiscriminate brown blob. Given that our interest is in locating the famous façade, it seems true to say that I see it, even though I can't make out any surface details. On the other hand, there are contexts in which it seems true to say that I don't see the façade: for example, one in which the kind of wood composing the façade is at issue.

What these kinds of cases show is that the context-sensitivity of seeing-ascriptions runs deeper than our initial formulation, (CS), allowed. Context does not just determine

whether the subject counts as seeing a given object, over and above its facing surface: it also determines whether she counts as seeing its facing surface in the first place.

#### *2.3.4 More evidence for context-sensitivity*

So far, we have been considering cases in which the truth-value of “*S* sees *O*” seems to be varying with the interests of ascribers. But the evidence for contextualism is not just limited to cases. For notice that if ‘sees’ really is context-sensitive then speakers should be able to make the relativity to specific interests explicit when required. For instance, take the context-sensitive adjective ‘flat.’ As Hawthorne (2004) has pointed out, if I say “That field is flat” and you challenge me by pointing out that the field has small holes in it, I will clarify—rather than retract my statement—by saying something along the lines of “What I was saying is that it’s flat *for a football field.*”

If we examine the linguistic devices that an ascriber has at her disposal whenever her seeing-ascriptions are challenged, it is clear that the contextually determined standards can be made explicit. For instance, suppose we are walking towards the Taj Mahal. I complain that I cannot see it yet. You challenge me by pointing to one of its spires and uttering, “Yes you can. That’s one of its spires!” In clarifying, I will say something like, “Yes, I realize that I can see one of its spires. What I was saying is that I don’t see enough of the Taj Mahal in order to appreciate its architecture.”

In addition to providing further support for the view that ‘sees’ is context-sensitive, the fact that we clarify seeing-ascriptions in the manner described above also undermines an obvious worry. According to this worry, when we deny that a subject sees an object (over and above its facing portion), we are just speaking loosely. For instance, in denying that I see the Taj Mahal (over and above one of its spires), I’m just speaking loosely. Strictly speaking, I do see the Taj Mahal because seeing part of an object is al-

ways sufficient for seeing the object itself.<sup>39</sup>

The problem with the loose-talk approach is that it fails to take into account how we actually respond when our seeing-ascriptions are challenged. As was just pointed out, if my primary purpose is to appreciate the architecture of the Taj Mahal, and the only part that is not occluded is one of its spires, I will clarify by making my interests as ascriber explicit (in the way described above). What I will not say is that I was just speaking loosely. For instance, I will not say something analogous to what I would say if you challenged my utterance of “It’s 12:30pm” by pointing out that it is, in fact, 12:33pm. In reacting to your challenge I will say something like, “Well, yes, I was speaking loosely. I realize that the exact time is 12:33pm.”

The fact that we clarify seeing-ascriptions in the way described above also undermines the worry that my utterance of “I can’t see the Taj Mahal” is false in all contexts, but that it *seems* true (in the given context) because I’m conveying a true implicature. What might that implicature be? In general, true implicatures are only fit to explain the appropriateness of a false utterance in cases of hyperbolae, irony, metaphor, or loose talk. For instance, it is appropriate to utter the false sentence “Smith is as fast as lightning” if the conversational context concerns whether Smith is a fast runner, because it implicates the true proposition that Smith is a very fast runner indeed. But in uttering the sentence “I can’t see the Taj Mahal,” I do not seem to be engaging in this kind of speech act at all—to put it in Gricean terms, I don’t seem to be flouting any conversational maxims.<sup>40</sup> This is reinforced by the fact that when my assertion is challenged, I will clarify the proposition expressed by making my interests explicit: what I will not do

---

<sup>39</sup> Recall that Broad (1925, 149) and Moore (1953, 34) hold the opposite view: according to them, when the folk ascribe states of seeing ordinary 3D objects (over and above their facing surfaces) they are speaking loosely. Problems exactly analogous to those raised below apply to this version of the loose-talk approach.

<sup>40</sup> See Grice (1975) for a discussion of conversational maxims and the generation of implicatures.

is concede that I was being hyperbolic or speaking metaphorically.<sup>41</sup>

### 2.3.5 *Subject-sensitive invariantism?*

I have been arguing that which head-world relation ‘sees’ picks out varies with the interests of ascribers. The view I’m defending is thus clearly analogous to standard versions of epistemic contextualism, according to which the extension of ‘knows’ varies with the interests of ascribers—there is no such thing as *the* knowing relation.

Now, one of the epistemic contextualist’s key opponents is the subject-sensitive invariantist. The subject-sensitive invariantist denies that the truth-value of “*S* knows that *p*” varies with the interests of utterers; however, she accepts that it varies with the interests of the subject, *S*. To put it another way, the subject-sensitive invariantist accepts that there is a unique knowing relation: it is just that she construes it as the relation that the subject, *S*, stands in with respect to a proposition just in case *S* meets the epistemic standard operative in her practical situation (not the utterer’s). In contrast, the epistemic contextualist claims that there are many knowing relations; ‘knows’ expressing different ones depending on the interests of the utterer.<sup>42</sup>

This invites an obvious question: why not adopt subject-sensitive invariantism about ‘sees’?<sup>43</sup> More specifically, why not adopt the view that there is a unique seeing relation: namely, the relation that a subject stands in with respect to an object just in case her interests are of the requisite sort (e.g. she is able to satisfy her interest in locating the object)? I think there are a number of reasons to favor contextualism about ‘sees’ over the subject-sensitive alternative.

---

<sup>41</sup> See DeRose (2009, 112–17) for a similar point in defense of the claim that the apparent context-sensitivity of ‘knows’ cannot be explained away in terms of implicatures.

<sup>42</sup> Proponents of subject-sensitive invariantism include Hawthorne (2004) and Stanley (2005). The view is sometimes referred to as interest-relative invariantism.

<sup>43</sup> Thanks to Gary Ostertag for pushing me on this issue.



First and foremost, consider the initial cases that we used to motivate contextualism. For instance, take the scenario in which you are waiting—very impatiently—for Frida the cow to be evaluated by the judge. Given *your* interest in Frida’s being judged as soon as possible—coupled with the fact that the only part of Frida that is unoccluded from the judge’s vantage point is her tail—you seem to speak truly in uttering, “Ugh! The judge can’t see Frida at the moment: she can only see her tail!” On the other hand, the judge’s sole interest is in making sure that all of the cows are in the arena: she is not concerned with the identity and location of any specific cow just yet. The judge thus seems to speak truly in uttering, “I see all 10 cows now.”

The contextualist has a very straightforward explanation of the fact that both utterances seem true: they both seem true because they *are* both true—the truth-conditions of seeing-ascriptions being sensitive to the interests of the ascriber. The subject-sensitive invariantist, however, needs to tell a story about why your utterance is false, even though it seems true. And as we have already seen, the prospects for telling a story of this sort are dim—as I argued above, in cases of this sort, you certainly don’t seem to be engaging in hyperbolae, metaphor, irony, or loose talk.

In order to bolster this last point, consider how ordinary speakers go about ascribing states of object-seeing when ignorant of the subject’s interests. For example, suppose we are sitting next to one another on the bus, both looking out the window. From our common vantage point, the Empire State Building’s spire is the only part of it that is not occluded by other buildings. Now, suppose you don’t know me. Moreover, you are completely ambivalent about whether I’m a tourist or a local; whether I’m trying to locate the Empire State Building; whether I’m trying to appreciate its architecture; and so on. All you know is that its spire is visible to me, and that I’m looking directly at it. In discerning what I see, does your ignorance about my interests cause you to hesitate? For instance, will you hedge and say something like, “He sees the spire of the Empire

State Building, but I'm not sure whether he sees the building itself"? Presumably, you will not. If what is at issue for *you* is the location of the building, you will say that in seeing the spire, I see the building itself.

Moreover, suppose I were to remove your ignorance by informing you that I'm gazing in the direction of the Empire State Building because I want to discern its architectural design, the materials it was made with, and so on. I therefore insist that I cannot see the Empire State Building itself: only its spire. In responding to this new information, you will not retract your statement: rather, you will employ the clarification technique discussed above. That is to say, you will clarify your initial statement by making your interests explicit: e.g. you might say, "All I was saying is that you see enough of the building to locate it." In other words, you will utilize the same kind of clarification technique that you do when, for example, your flatness-ascription is challenged by someone who harbors interests that diverge from yours (e.g. you are a football player and your interlocutor is a surveyor).

Another problem with subject-sensitive invariantism concerns 'now you see it, now you don't' sentences. For instance, the subject-sensitive invariantist must commit to the truth of bizarre sentences, such as "He sees the Empire State Building now, but he won't as soon as I get him to wonder about its architecture"; "She sees the cat right now, but she won't once I ask her to identify its breed"; "Oliver sees me now, but if I can just get him to wonder about my weight, he won't be able to see me anymore." The subject-sensitive invariantist is committed to sentences of this kind being true because the subject's interests can obviously vary over time, thereby altering what she counts as seeing, even though her vantage point remains constant.

In contrast, the contextualist is *not* committed to the truth of 'now you see it, now you don't' sentences, because the interests that determine the truth-conditions of seeing-ascriptions concern the *current* interests of the utterer. For instance, if what's at issue

for *you* right now—as the utterer—is the location of the Empire State Building, then your utterance of the following is just false: “Ben sees the Empire State Building now, but he won’t see it as soon as I get him to wonder about its architecture.” This utterance is false because seeing the building’s spire suffices for locating it, and that is what your current interest concerns. It doesn’t matter if you *will* be interested in the building’s architecture five seconds after uttering the sentence above: the interest that attended, and had a causal role to play in generating, your utterance of the sentence above solely concerned the building’s location.<sup>44</sup>

Finally, notice that even if the truth-values of seeing-ascriptions vary with the interests of the subject, not the interests of the ascriber, this does not undermine my claim that there is no such thing as *the* seeing relation. The reason is that seeing-subjects often harbor interests that pull in different directions. For instance, suppose I want to locate the Empire State Building *and* appreciate its architecture. Given my vantage point, the first interest is being satisfied, whereas, the second is not. If we are subject-sensitive invariantists, what are we to make of this kind of situation?

According to the subject-sensitive invariantist, seeing is a three-placed relation whose argument-places are occupied by a subject, an object, and an interest. But in the case just described—which is far from being an exotic one—we have different interests generating different verdicts on what I qualify as seeing. Should we conclude that your statement of “Ben sees the Empire State Building” is neither true nor false? I don’t think this would necessarily be a problematic conclusion for the subject-sensitive invariantist to draw (no doubt, some will find it troubling). The point I want to make here is

---

<sup>44</sup> See DeRose (2009, 204–6) for a persuasive argument that the epistemic contextualist is not committed to the truth of analogous ‘now you know it, now you don’t’ sentences. His arguments carry straight over to the case of ‘sees.’ See Schaffer (2006) for some other arguments against (epistemic) subject-sensitive invariantism.

just that to draw this conclusion would be to relinquish the claim that there is a unique seeing relation: rather, what we would have are various equally acceptable candidates. For the contextualist, the ascriber's interests determine which candidate is expressed by her utterance of 'sees', whereas, for the subject-sensitive invariantist, 'sees' is just indeterminate between the various candidates. Either way, we must discard the prejudice that there is a privileged seeing relation, and that is an important result.

### *2.3.6 Looking ahead*

So there are compelling reasons for thinking that the folk deploy a context-sensitive notion of object-seeing. Even still, I suspect that some readers will harbor the following kind of worry: "Your contextualist theory is not a theory of seeing per se: it is a piece of ordinary language philosophy, and as such, it merely tells us how the folk go around ascribing states of seeing. If we want to home in on a theoretically interesting notion of object-seeing, we should examine the various explanatory roles that, as theorists, we would want it to play. What's more, there is no guarantee that the notion (or notions) we uncover will be context-sensitive."

In what follows, I address this worry head-on by focusing on the explanatory roles of object-seeing. First, I assess explanatory roles that philosophers of perception have traditionally been concerned with. I then move on to an assessment of the notion of object-seeing that earns its explanatory keep in vision science. In both cases, I argue that there are good reasons for thinking that the notions of object-seeing that emerge are context-sensitive ones.

## **2.4 Contextualism and the anchoring of *de re* thoughts**

One of the fundamental roles of object-seeing is to put us in contact with objects in the environment. For example, upon seeing the facing portion of a crocodile—which hap-

pens to be its tail—suppose Roger acquires the thought, *THAT IS A CROCODILE*. Presumably, the demonstrative element of Roger’s thought refers to the crocodile, not its tail. What makes this the case? This is obviously a difficult question, but it is generally agreed that seeing an object partly explains the capacity to form perceptual demonstrative thoughts about it. Is a context-sensitive notion of object-seeing fit to play this explanatory role?

#### *2.4.1 Siegel’s regimented notion of object-seeing*

Siegel (2006) seems to think not. According to her, if we are to home in on a notion of object-seeing that plays the explanatory role described above, ordinary language reports must be regimented, for they are “not exact guides to the sort of object-seeing at issue” (2006, 431). More specifically, they must be regimented in a way that removes any context-sensitivity. In order to illustrate her point, Siegel provides the following example (2006, 432).

Franco and Ray are at the docks, standing by a boat that is completely wrapped in a tarp. Someone informs them that it is the renowned Lady Windermere. Later that day, someone asks Franco about the color of the Lady Windermere, to which he replies, “I don’t know. I didn’t actually see her.” On the other hand, suppose Franco is asked whether the Lady Windermere has docked yet, to which he replies, “Yes. I saw her by the dock.”

Both of Franco’s utterances seem true. However, according to Siegel, this does not mean that the notion of object-seeing that explains our capacity for perceptual demonstrative thought is context-sensitive. On the one hand, we could deny that Franco speaks truly on both occasions, on the grounds that “our best theories of object-seeing overrule our intuitive judgments about when ordinary utterances are true” (Siegel 2006, 432). In other words, we could reject contextualism about the ordinary notion of object-

seeing on the grounds that a context-sensitive notion is not fit to play (at least one of) the explanatory roles that we would want it to as theorists. Call this *the revisionist view*. On the other hand, we could

... go to the other extreme and hold that the only theoretically useful notion of object-seeing is one that is tracked exactly by ordinary language reports—and hence that whether S sees *o* depends on more than the factors held constant in the two examples given in the text. This second option is a threat to the option I favor only if the more restricted notion that I've zeroed in on is not a legitimate subject-matter for theorizing. But far from seeming illegitimate, the subject-matter I've defined is one of the central explananda in the theory of intentionality. (2006, 432)

Call this *the ambiguity view*, for it says that the notion of object-seeing that explains reference-fixing for perceptual demonstrative thoughts is distinct from the ordinary context-sensitive one. Does the ambiguity view have anything going for it? What about the revisionist view? In what follows, I argue against both.

#### *2.4.2 Contextualism and reference-fixing for demonstrative thoughts*

Why think that in telling a reference-fixing story for perceptual demonstrative thoughts, only a context-*invariant* notion of object-seeing will fit the bill? Perhaps the worry is that if the referents of my thoughts are determined by which objects I count as seeing, and which objects I count as seeing is a context-sensitive affair, it follows that which objects my thoughts count as referring to is thereby a context-sensitive affair as well. But is this inference a good one? For instance, if whether Roger counts as seeing the crocodile (over and above its tail) varies with the interests of ascribers, does it follow that whether his thought refers to the crocodile varies with the interests of ascribers

too?

Arguably, it would be problematic if contextualism about object-seeing had this consequence. Certainly, the truth-conditions of Roger's thought—*THAT IS A CROCODILE*—seem to concern the crocodile, not its tail. And it is no doubt counterintuitive to claim that my utterance of “Roger's thought refers to *the crocodile's tail*” is true relative to my context of utterance: if that were the case then Roger's thought would count as false (relative to this context) on account of attributing the property of *being a crocodile* to the crocodile's tail. Moreover, Roger certainly does not behave as if the tail is a crocodile: e.g., if we were to detach the tail and place it in front of him, he wouldn't run away in fear of being bitten!

Fortunately, as contextualists about the truth-conditions of seeing-ascriptions, we can be invariantists about the truth-conditions of perceptual thoughts by appealing to a context-*insensitive* reference-fixing mechanism. That is to say, the invariantist about thought content can posit a mechanism that takes a visual experience as input—where that experience puts the subject into different ‘seeing’ relations relative to different contexts—and produces a perceptual thought with determinate, context-invariant, referents. How might a mechanism like this work?

An obvious option is to claim that Roger's thought picks out the crocodile—not its tail—because it contains a complex demonstrative of the appropriate kind. For example, perhaps Roger's thought is of the form, *THAT ANIMAL IS A CROCODILE*, where *being an animal* is a property instantiated by the crocodile, not its tail.<sup>45</sup>

This is not the place to assess competing theories about what fixes the referents of perceptual demonstrative thoughts. The crucial point is that whichever mechanism we posit, there is no reason why it cannot fix the referents of perceptual demonstrative thoughts in virtue of context-*invariant* facts about the subject and her environment. For

---

<sup>45</sup> See Strawson (1959) for a version of this view.

instance, if we go ahead and provide an account that involves an appeal to complex demonstratives, then we can give a reference-fixing story that appeals to the ascriber-*independent* facts concerning which properties are represented by the complex demonstrative in question (e.g. the property of *being an animal*).<sup>46, 47</sup>

### 2.4.3 Contextualism and the differentiation constraint

So being contextualists about object-seeing does not necessarily undermine our ability to explain reference-fixing for perceptual demonstrative thoughts. But there are also positive reasons for thinking that the notion of object-seeing that does this explanatory work is a context-sensitive one.

In order to see why, consider Siegel's own account of the notion of object-seeing that explains reference-fixing for perceptual demonstrative thoughts. Following Dretske (1969, 18–35), she defends the following condition:

*Differentiation condition:* If  $S$  sees  $o$ , then  $S$ 's visual phenomenology differentiates  $o$  from its immediate surroundings. (2006, 434)

Siegel motivates the differentiation condition with examples in which the subject is not in a position to form *de re* perceptual thoughts about an object because it has not been discriminated from its immediate surroundings by her visual experience. She describes a

---

<sup>46</sup> There may well be reasons for thinking that which *properties* one represents (in thought or perception) is a context-sensitive affair. However, it is extremely plausible that the demonstrative element of Roger's thought refers to the crocodile, not its tail. Claiming that which *property* his thought counts as representing is a context-sensitive affair does nothing to alter that point. Thus, in order to avoid the worries mentioned above we still need to posit a context-invariant reference-fixing mechanism for his thought's demonstrative element.

<sup>47</sup> For some non-descriptivist reference-fixing mechanisms that are compatible with the claim that we refer, in thought, to the objects we perceive in a context-*invariant* fashion, see Campbell (2002) and Dickie (2011).



case in which your friend, Franco, is doing stunts in the sky by suspending himself from invisible fibers. At present, he is dressed in red and so you can easily differentiate and thereby form *de re* thoughts about him (e.g. *THAT IS FRANCO IN THE SKY*).

On the other hand, suppose Franco were to paint himself blue, so as to blend in perfectly with the sky. Another subject, *S*, looks towards Franco. However, given that Franco is visually indistinguishable from his immediate surroundings, *S*'s experience does not put her in a position to form *de re* beliefs about him. As Siegel puts it, "S does not see Franco in the sense that matters, given the theoretical purpose of the notion of object-seeing that is at issue" (2006, 434).

Further support for the differentiation condition comes from a consideration of the action-guiding role of vision. As we saw in chapter 1, other than putting us into cognitive contact with objects, another role of object-seeing is to put us in a position to perform actions on them—actions such as grasping, pushing, lifting, and so on. Unless seeing the cup on the table requires that I have visually differentiated it from its immediate surroundings, it is unclear how seeing it could explain my successfully picking it up.

One might press against the thesis that seeing *O* requires differentiating it from its immediate surroundings by pointing out that an undifferentiated object can still have an effect on the phenomenology of one's overall experience. For instance, even though *S* does not differentiate Franco from his immediate surroundings, Franco is still (partly) causally responsible for the fact that *S* has a visual experience as of a uniformly blue sky. The problem with this view, though, is that it merely gets at the notion of *property-seeing*: *S* clearly sees the *blueness* that Franco instantiates, but that does not entail that she sees Franco himself. What explanatory payoff would we enjoy by insisting that, in addition to the blueness, *S* sees Franco himself?

Siegel's differentiation condition is thus compelling, but is the notion of object-seeing that emerges a context-*invariant* one? In fact, there are two reasons why the notion that

emerges is likely context-sensitive. First, as Dretske has pointed out, the notion of differentiation is likely context-sensitive itself:

How much of  $D$  must be visually differentiated? How much of  $D$  must look some way to  $S$ ? ... In general, I think it is clear that one need not see every part of something to see it, nor need one see all of its surface or even all of one surface. How much of a thing  $S$  must visually differentiate in order to see <sub>$n$</sub>  it is a question that, in the abstract, divorced from the sort of thing he is seeing, and the circumstances under which he sees it, cannot be answered. (1969, 27)

In clarifying what he means by this, Dretske goes on to say the following:

The conversational context and one's particular interests (where is it? what is it? is it a so-and-so?) will also affect the question of whether enough of  $D$  was seen to see  $D$ . (1969, 28)

Of course, given the heterogeneity of ascribers' interests (discussed above), we need to qualify Dretske's remarks a little here, for the issue is not whether *enough* of  $D$  has been differentiated in order for it to qualify as being seen; rather, the issue is whether the right *kind* of part has been differentiated. In any case, the important point is that in addressing the question of which part of an object must be differentiated in order for the object itself to count as having been differentiated, we arrive back at the exact kind of question we started with. And the answer is exactly analogous to the one we gave: that is to say, whether differentiating part of  $O$  suffices for differentiating  $O$  itself is a

context-sensitive affair.<sup>48</sup>

How might the invariantist respond at this point? Might it be that the notion of visual differentiation in play is a regimented context-invariant one? One way of trying to substantiate this claim would be to cash out the notion of visual differentiation in representational terms. For instance, perhaps visually differentiating *O* from its immediate surroundings amounts to tokening property-representations that are satisfied by *O*, but not its facing portion.

For instance, go back to the scenario in which, from Roger's perspective, the crocodile's tail is its visible portion. Call the 3D shape of the tail, *T*; and call the 3D shape of the crocodile, *C*. Perhaps visually differentiating the crocodile's tail from its immediate surroundings requires visually representing the property of *being T-shaped*; while differentiating the crocodile itself requires visually representing the property of *being C-shaped*. If Roger only manages to represent the first property, he only sees the tail; whereas, if he manages to represent the second, he sees the crocodile itself. If we take on the plausible assumption that whether Roger's visual system represents the property of *being C-shaped*, the property of *being T-shaped*, or both, is a context-invariant affair then it seems to follow that which objects he has visually differentiated is a context-invariant affair as well.<sup>49</sup>

---

<sup>48</sup> There are, in fact, two ways for context-sensitivity to get a foothold here. First, given the arguments provided above, it is probably a context-sensitive affair whether visually differentiating a given part of *O* suffices to qualify as having seen *O* itself. For instance, does differentiating your hand from the couch (which you are hiding behind) suffice for me to count as seeing you? If what I have argued above is right, that will depend on the interests of the ascriber. Second, it is probably a context-sensitive affair whether differentiating a given part of *O* suffices to count as having differentiated *O* itself: that is to say, the notion of *differentiation* is context-sensitive.

<sup>49</sup> Once again, I will not argue for the view that which properties are represented in perception is a context-invariant affair. In brief, though, one reason for denying that it is a context-sensitive affair is that the phenomenal character of a perceptual experience is plausibly constituted by those facts concerning which properties it represents. Given that the phenomenal character of my experience does not seem to vary with the interests of ascribers, this suggests that

The problem with this proposal, though, is that one can see an object even though one's experience is illusory with respect to its color, shape, size, location, kind, and so on. In the right conditions, I might see the white house on the horizon as a small red disc; the 2D photo on the wall as a 3D person; the person over my shoulder as directly in front of me (where she is being reflected in the mirror); and so on. Visually misrepresenting the shape, size, color, kind and position of an object does not remove my capacity to form *de re* thoughts about it: it is just that these thoughts will be inaccurate.

Are there any properties that I cannot get wrong if I'm to qualify as seeing the object in question? Plausibly, my visual system must distinguish figure from ground, and represent the presence of a cohesive, bounded, and spatiotemporally continuous object.<sup>50</sup> Unless seeing *O* requires that my visual system has effected these segmentation and grouping procedures, it would be mysterious how seeing *O* could put me in a position to form *de re* thoughts about it; it would also be mysterious how seeing *O* could put me in a position to perform successful actions on it.

Importantly, though, as I argued in chapter 1, it is doubtful that appealing to segmentation and grouping processes will give us context-invariant notions of visual differentiation and seeing. The reason is that both 3D objects and their facing portions count as cohesive, bounded, and spatiotemporally continuous particulars; therefore, which ones qualify as having been differentiated via processes of segmentation and grouping will likely be a context-sensitive affair. For instance, suppose I'm watching a car as it moves

---

which properties my experience represents does not vary in this way either. See Byrne (2001) for a discussion (and defense) of the view—often referred to as *intentionalism*—that the phenomenal character of a perceptual experience supervenes on its representational content. Causal-informational theories (e.g. Fodor 1990, Prinz 2002), teleological theories (e.g. Millikan 1984, 2004), and quality-space theories (e.g. Rosenthal 1991, 1999, 2005, 2010, in press) all explain how properties are represented in perception by appealing to ascriber-independent facts about the perceiver and her environment.

<sup>50</sup> This appears to be Burge's view of seeing (2009, 31; 2010, 456).

along the horizon. Both the car and its facing surface are cohesive, bounded, and spatio-temporally continuous particulars, and both are plausibly causes of my visual experience. Thus in order for my visual experience to qualify as having differentiated either entity from its immediate surroundings, something more is needed. But what other resources do we have? If we appeal to representations of color, shape, size, kind or location then we face a familiar problem: one can see an object, and thereby be in a position to form *de re* thoughts about it, even if one's visual experience gets these features very wrong. If, on the other hand, we embrace contextualism about the notion of visual differentiation, the problem does not arise in the first place. The reason is that if contextualism is true then whether I have visually differentiated the car (or its facing surface) from its immediate surroundings will simply vary with the interests of ascribers. There is thus no need to search for elusive constraints on the notion of visual differentiation in play: constraints that would yield a unique head-world relation.

## **2.5 Contextualism and vision science**

In assessing whether a theoretically fruitful notion of object-seeing is context-sensitive, so far we have examined capacities that are manifested downstream of vision. But what about those decidedly *visual* capacities that vision scientists are in the business of explaining? Do they permit context-sensitivity, or, is the notion of object-seeing that falls out of their explanations a context-invariant one? If the latter turns out to be the case then, once again, there are two options. According to the revisionist view, the conclusion to draw would be that the folk commitment to context-sensitivity is mistaken. According to the ambiguity view, the conclusion to draw would be that the folk notion of seeing and the one that earns its keep in vision science are just different notions altogether.

### 2.5.1 Problems with the revisionist view

The view that in order to ascertain what seeing is like we should uncover the notion that earns its keep in vision science figures prominently in the work of theorists such as Pylyshyn (2003), Burge (2010) and Block (2013, forthcoming). For instance, Block (2013, 180) claims that “in discussing seeing we should be focused on cutting nature at its joints.” But is the folk notion of object-seeing beholden to findings in vision science? More importantly, if it turns out that the notion of seeing deployed in vision science is context-*invariant*, will we have thereby found out that the folk commitment to context-sensitivity is mistaken? I highly doubt it.

First, notice that if we are contextualists about seeing-ascriptions then that does not commit us to any special constraints on the internal vehicles that carry visual experiences. In fact, the contextualist and the invariantist can agree on the nature of those internal states that put the subject into seeing relations with environmental objects. For instance, they can agree on whether visual states are discrete and symbolic (in accordance with classical computational approaches to vision) or distributed (as proponents of connectionist models would have it).

In general, it is very hard to see how findings in vision science could be marshaled in support of the claim that the ordinary notion of seeing expresses a unique head-world relation (irrespective of context). Consider two of the relations that, according to the contextualist, are expressed by ‘sees’ (relative to different contexts). There is the relation that only holds between a subject’s visual experience and *the facing surfaces* of objects. On the other hand, there is the relation that holds between a subject’s visual experience and the ordinary objects of which those facing surfaces are parts.

The dispute between the contextualist and the invariantist does not concern whether these two relations do, in fact, hold between a normal perceiver’s experiences and objects in the world. Rather, the dispute concerns whether one, and only one, of these re-

lations is expressed by the verb ‘sees,’ or, whether it varies according to the context. That being the case, if it turns out that explanations in vision science only invoke one of these head-world relations then, at best, that would support the view that the notion of object-seeing deployed in vision science is context-*invariant*. But that still leaves it completely open whether the folk notion is context-sensitive. In other words, if only one of the head-world relations described above earns its explanatory keep in vision science then that would merely provide a reason to adopt the ambiguity view. In order to support the revisionist view, the contextualist would have to be committing to substantive, and erroneous, claims about the nature of the perceptual vehicles and processes that underlie our visual experiences. And as I argued above, the contextualist takes on no such commitments.

Might it be that the folk notion of object-seeing is a natural kind concept, and as such, its extension is fixed in a way that is independent of human interests? On this view, *object-seeing* would be akin to other paradigmatic natural kind concepts (e.g. *water*), whose extensions are widely seen as being fixed by external relations to local samples: relations that obtain independently of the interests of humans.<sup>51</sup> But what is the independent motivation for this kind of view? I have argued that the linguistic data support contextualism, and as such, any view according to which the extension of the folk notion is fixed in a way that is independent of human interests will fall short of accommodating the data. Even if we suppose there to be a unique head-world relation that is privileged by the explanations of vision scientists, this does not by itself settle the question of whether the folk notion picks out the same relation.

### 2.5.2 *The ambiguity view*

How might one argue for the view that, unlike the folk notion of seeing, the one de-

---

<sup>51</sup> See Putnam (1975) and Kripke (1980).

ployed in vision science is context-*invariant*? In what follows, I will not argue against the ambiguity view per se, for it may well be that the folk notion of object-seeing and the one that earns its keep in vision science are both context-sensitive, even though they are, indeed, distinct notions (e.g. because they diverge in some of the explanatory roles they play). My aim will be the more modest one of arguing that the scientific notion is context-sensitive, regardless of whether it is the same exact notion as the folk one.

## 2.6 Arguments for the context-sensitivity of the scientific notion

Among the various head-world relations that we have discussed, is one of them uniquely suited to play the role of *the* seeing relation that earns its keep in vision science? I highly doubt it. More strongly, I think there are compelling reasons for thinking that which head-world relation the theorist invokes will depend on her explanatory interests.

### 2.6.1 Individuating biological traits and individuating states of seeing

In order to motivate the view that the vision scientist's explanatory interests determines which 'seeing' relation she is invoking, I want to start by briefly considering the individuation of biological traits, for it is very plausible that: (i) this depends on the explanatory context of the theorist, and (ii) similar considerations apply to the individuation of states of seeing.

There are three main ways to type biological traits:

- (1) *Functional criteria*: an object belongs to a given trait type just in case it performs some function, *F*. For instance, the function of a lung is two-fold: it transfers oxygen from the atmosphere into the bloodstream, and it releases carbon dioxide from the bloodstream into the atmosphere. Something is typed as a lung only if it performs this two-fold function.



- (2) *Morphological criteria*: an object belongs to a given trait type just in case it has the requisite morphological properties. For instance, an object is a lung just in case it is the right shape, color, and size.
  
- (3) *Homological criteria*: two objects belong to the same trait type just in case they have a shared ancestry. For instance, the lungs of *homo sapiens* and gorillas belong to the same type because they have a shared ancestry.

The important point here is that token biological traits are typed differently depending on the explanatory task at hand. As an example, consider bats and birds. Their wings perform the same function, and their shapes are similar in relevant respects. If we were to individuate biological traits based solely on functional and morphological criteria, the wings of bats and birds would therefore qualify as belonging to the same type. However, the wings of bats and the wings of birds are not homologues: they have a different evolutionary history. Thus if we were to individuate traits solely in terms of homological criteria, the wings of bats and birds would be seen as belonging to different types.

Given that there are different ways to type biological traits, it is widely held—in both biology and the philosophy of biology—that we should use different criteria, depending on the explanatory context. If we are interested in explaining how birds and bats manage to propel themselves through the air then typing their wings in terms of functional criteria makes good sense. On the other hand, if we are interested in explaining the specific differences in morphology between the wings of birds and the wings of bats, homological criteria will be central. The key point is that the explanatory interests of the theorist determine how the entity in question is to be typed: there is just no in-

terest-independent way to do it.<sup>52</sup>

Now, as Nanay (2015) argues, if the individuation of biological traits is sensitive to the explanatory context, why shouldn't this also be the case for the individuation of perceptual states?

If the individuation of other biological traits depends on the explanatory project, we should expect that so does the individuation of perceptual states (see Matthen 1998 for a similar point). Here is why: perceptual states are states in the perceptual system, that is, in an evolved biological mechanism. And as the individuation of the states of other evolved biological mechanisms, like the *systole* and *diastole* states of the heart, the individuation of the states of our perceptual system is also sensitive to the explanatory project at hand. If we have good reasons to doubt that there is one and only one way of individuating the *systole* state of the heart, we also have good reason to doubt that there is one and only one way of individuating perceptual states. (2015, 327)

In arguing that token perceptual states are individuated in a way that is sensitive to the explanatory interests of the theorist, Nanay focuses on the dispute between representationalists (those who claim that perceptual states are individuated in terms their representational contents) and relationists (those anti-representationalists who think that perceptual states are to be individuated in terms of those objects that they put us into the perceiving relation with). Of course, that particular debate is not my concern in this

---

<sup>52</sup> See Nanay (2010b, 2011, 2012, 2015) for a defense of this view. For a context-insensitive approach to the individuation of *species*, see Devitt (2008, 2009, 2011). See Stanford (1995) for a defense of pluralism about species individuation.

chapter.<sup>53</sup> However, notice that if the line of reasoning given above is correct, there is no reason why we should not extend it to the individuation of *states of seeing*. That is to say, why expect there to be a unique way to type head-world relations into seeing and non-seeing varieties? There are a host of different head-world relations that we can appeal to in specifying which objects a given subject counts as seeing in manifesting her (various) visual capacities: which one the theorist picks out will be determined by whichever interests are engendered by the specific capacity she is seeking to explain.

Now, I do not take this to be anything like a dispositive argument. Nonetheless, I do think that it provides prima facie support for the view that the notion of object-seeing deployed in vision science is a context-sensitive one. In any case, I want to bolster the argument now by focusing on the visual object-tracking system: an evolved mechanism that we share with many nonhuman animals. I will argue that which head-world relation the theorist invokes in explaining successful instances of object tracking will depend on her explanatory interests, just as the considerations given above predict.

### *2.6.2 MOT and explanatory interests*

As was explained in chapter 1, Pylyshyn (2001, 2003, 2006, 2007) has argued forcefully that ordinary subjects track objects via the assignment of visual indexes in early vision. For present purposes, the important point is that Pylyshyn regards the regimented notion of object-seeing that falls out of his account of multiple-object tracking as a theoretically fruitful one, even if it departs from the ordinary notion of object-seeing which, according to him, fails to “cut nature at her joints” (2003, 51).

It is certainly plausible to construe the relations that hold between visual indexes and tracked objects as object-seeing ones. As was argued above, one of the core features of object-seeing is that one can see an object without representing its color, shape, size,

---

<sup>53</sup> In chapter 3, I take up the related issue of whether perceptual content is object-dependent.

kind or location accurately. However, one must visually discriminate an object from its immediate surroundings in order to qualify as seeing it: lest we end up with an explanatorily vacuous notion of object-seeing. Fortunately, the notion of object-seeing that falls out of Pylyshyn's account of MOT meets these criteria: one can track an object without getting its color-, shape-, size-, and kind-properties right; and tracking an object requires that early vision has discriminated it from its immediate surroundings (I elaborate on this issue in more detail below).<sup>54</sup>

For these reasons, I agree with Pylyshyn that there is a theoretically fruitful notion of object-seeing according to which one sees an object just in case one's visual system has assigned a visual index to it. But does it follow that this notion is context-invariant (unlike the ordinary notion)? I highly doubt it.

In order to see why, notice that—as with any causal or informational theory of reference—there are several causal stories to tell about which environmental particulars are referred to by visual indexes. Suppose I'm tracking a car that is passing me by. My visual system maintains causal contact with the car itself; with its facing surface; with its undetached parts; and so on—that was the moral of chapter 1. Of course, we could try to privilege one (or some combination) of these candidates by putting constraints on which causal relations (holding between visual indexes and tracked objects) are the reference-conferring ones. Needless to say, the track record of such attempts is pretty dis-

---

<sup>54</sup> As far as I'm aware, Pylyshyn does not explicitly discuss cases in which the subject tracks an object, despite the fact that its location is visually misrepresented. Clearly, though, cases of this sort do occur. For instance, just consider a scenario in which you are tracking a fish from a boat: your visual experience misrepresents the exact location of the fish, due to the distorting effect of the water, but you still keep track of it as it moves.

mal, and as I have argued in chapter 1, they are not motivated by relevant empirical data.<sup>55</sup>

Perhaps we could avoid these causal indeterminacies—without relinquishing the view that visual indexes track objects without encoding their colors, shapes, sizes, locations or kinds—by abandoning a purely causal theory and adopting the kind of view (discussed above), according to which an object is only seen if it has been visually differentiated from its immediate surroundings. In the context of Pylyshyn’s visual indexing view, the analogous claim would be that a *visual index* picks out an object (partly) in virtue of the fact that it has been differentiated from its immediate surroundings. In fact, in some places, Pylyshyn seems to endorse this very view:

- (1) early visual processes segment the visual field into feature-clusters ... and (2) recently activated clusters compete for a pool of four to five visual indexes (2001, 146)

Elsewhere, Pylyshyn states that MOT

operationalizes the notion of ‘primitive visual object’ as whatever allows preconceptual selection and MOT. Note that objecthood and object-identity are thus defined in terms of an empirically established mechanism in the human early vision system. A certain (possibly smooth) sequence of object locations will count as the movement of a single visual object if the early vision system groups it in this way—i.e. if it is so perceived. (2001, 143–4)

---

<sup>55</sup> Grice (1961) is the *locus classicus* of causal theories of perception. See Gates (1996) for a compelling argument that that no purely informational-causal account of reference (be it perceptual or cognitive) can avoid the kind of indeterminacy described above.

In other words, the conception of objecthood operative in explanations of MOT is a thin one: all that is required for a particular to count as an “object”—in this *thin* sense of term—is for it to be a bounded, cohesive, particular that traces out a continuous path. Ordinary 3D bodies and their facing surfaces both count as objects in this sense of the term, and thus there is nothing to choose between them as far as fixing the reference of visual indexes goes.<sup>56</sup>

We are thus back where we were above, armed with a differentiation condition that does not determinately distinguish between 3D objects and their facing surfaces: both seem like equally good candidates. And if we try to adjust the differentiation condition by appealing to visual representations of color, size, shape, and so on, we face the problem discussed above: namely, one can see an object without one’s visual experience getting its color, shape, size, kind or location right. In any case, we have already seen that MOT proceeds independently of these kinds of representations.

For all these reasons, I think that it is probably a fool’s errand to seek out *the* seeing relation that figures in the selection and tracking of objects. There is just no such thing. Rather, what we have is a set of equally good candidates, and only pragmatic considerations will determine which one the theorist invokes on any given occasion. As an illustration, reconsider the case in which I’m watching a car as it travels along the horizon. What is it exactly that the relevant index picks out? Does it pick out the car itself? Does it only pick out its facing surface?

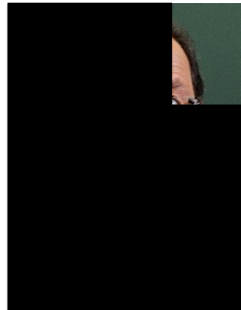
What I’m suggesting is that which option the theorist goes with will depend on her explanatory interests. If the project involves explaining *why* I’m tracking whatever it is that I’m tracking, and this involves an appeal to the fact that I recognized the object on the horizon *as a car*, it would be natural to construe the car itself—not its facing sur-

---

<sup>56</sup> See Clark (2006) and Dickie (2010) for similar points.

face—as the object picked out by my visual indexing system.<sup>57</sup> On the other hand, suppose the project involves explaining why I was unable to recognize the model of the car on the horizon—perhaps its facing surface was too nondescript for a match to be made with representations in long-term memory. In that case, it would be natural to construe me as only seeing, and tracking, its facing surface.

Similar remarks apply to explanations of amodal completion. For instance, suppose you were confronted with the scene depicted below in figure 9:



**Fig. 9** Partially occluded face

Consider the claim that you saw the partially occluded face, over and above its visible portion: call the relevant relation *seeing*<sub>1</sub>. Contrast this to the claim that you only saw the visible portion of the face: call the relevant relation *seeing*<sub>2</sub>.

If the theorist is concerned with facial recognition, and she wants to explain why it is that you were unable to recognize the man's face (it is François Hollande's), she will naturally construe you as having only seen *part* of his face, not the face itself (i.e. the theorist will invoke the *seeing*<sub>2</sub> relation). On the other hand, if she is interested in the amodal completion of partially occluded surfaces, and she wants to explain how it is that you visually represented the visible portion of Hollande's face *as* part of an oval-

---

<sup>57</sup> It is well known that top-down attention can drive the selection of targets. See Pylyshyn & Annan (2006) for a discussion.

shaped surface, she will naturally construe you as having seen the face itself (i.e. she will invoke the *seeing<sub>1</sub>* relation): whether you recognized this partially occluded surface *as Hollande's face* is not a relevant concern in this explanatory context.

## 2.7 Conclusion

Recall the question we started with: when is seeing part of an object enough to qualify as seeing the object itself? I have defended a contextualist answer to this question. Importantly, though, I have argued that context-sensitivity is not just a feature of the ordinary notion of object-seeing: it is a feature of the notion invoked in vision science, as well as the one invoked by philosophers of perception. In none of these domains can we talk about *the* seeing relation. Rather, which head-world relation the ascriber picks out will vary according to her interests. Heeding this fact is important, for if we fail to do so, we are in danger of searching for a privileged relation where there is none. Moreover, as explained in chapter 1, going contextualist allows us to explain the folk intuition that we often count as seeing ordinary 3D objects, despite the fact that it is indeterminate whether we visually represent them (as opposed to their facing portions).



# Chapter 3

## Object-seeing and the generalism versus particularism debate

### 3.1 Introduction

Imagine you are on a rickety bus. Your eyes are fixated on the man seated in front of you because the cup of coffee resting on his lap is about to get jolted off. Your experience of the cup's color, shape, and motion is thus extremely vivid. If we grant the widely held view that perceptual states are representational then your experience plausibly counts as representing the cup's color-, shape- and motion-properties. But what about the cup itself? Does your visual experience represent *it*?<sup>58</sup>

According to generalists, the content of a perceptual experience is completely independent of the facts concerning which objects, if any, the subject perceives. Moreover, generalists maintain that as long as *some* object satisfies the general content of the experience, it thereby counts as accurate. Particularists disagree, claiming that whether the subject's experience is accurate or not depends on whether she perceives any objects,

---

<sup>58</sup> Of course, not everyone accepts the claim that perceptual states are representational. See Travis (2004) for an argument against the representational view. For some recent defenses of the representational view, see Schellenberg (2011) and Nanay (2014).

and if so, how things are with those objects. In fact, particularists such as Tye (2009) and Schellenberg (2010, 2011) go even further, and construe the contents of perceptual experiences as being constituted by those objects, if any, that the subject perceives.

Rather than endorsing one side of the debate, I want to argue for a third option that has been neglected in the literature. According to the view I will argue for there is no fact of the matter as to whether the objects we perceive are among the contents of our perceptual experiences. What generalists and particularists are providing us with are equally acceptable ways of assigning contents to our perceptual experiences. In particular, I will argue that there is a systematic stalemate between generalists and particularists: whatever can be explained in terms an object-dependent content can be explained, equally as well, in terms of a purely general content, as long as the latter is supplemented by an appeal to a non-content-conferring perceiving relation.

Here is how I make my case. In sect. 2, I outline the varieties of generalism and particularism on offer. In sect. 3, I argue that generalists and particularists provide equally adequate explanations of phenomenology. In sect. 4, I argue that intuitions are unhelpful when it comes to determining who provides the best account of the accuracy conditions of perceptual experiences. Then, in sect. 5, I address the argument that we can only explain our capacity to demonstratively refer to objects in thought, as well as our capacity to acquire knowledge of those same objects, if we adopt a strong version of particularism. I argue that the generalist is, in fact, equally well placed to explain both capacities. In sect. 6, I defend the view that there is just no fact of the matter as to whether our perceptual experiences represent the objects we perceive. I then conclude by addressing two objections to the indeterminacy thesis. According to the first, we should favor generalism on grounds of parsimony. According to the second, we should posit two layers of content—an object-dependent layer and an object-independent one—rather than a single layer of indeterminate content. I argue that neither objection is

compelling. Finally, in sect. 7, I explain why it is indeterminate, rather than a context-sensitive affair, whether perceptual content is object-dependent.

### 3.2 The basic varieties of generalism and particularism

The key differences between the generalist and the particularist concern the ways in which they answer the following two questions: (1) Are the objects we perceive among the contents of our perceptual experiences? (2) Can we determine whether a perceptual experience is accurate or not without determining whether the subject perceives any objects?

#### 3.2.1 Generalism

The generalist answers “no” to the first question and “yes” to the second. She answers “no” to the first question because she construes perceptual content as purely general. As McGinn puts it, “when we are describing the content of an experience we should not make singular reference to the object of the experience” (1982, 51). One way to express McGinn’s point is to deny that the contents of perception are *object-dependent*, where a content is object-dependent just in case that very content could only occur given the existence of a specific object.<sup>59, 60</sup>

If the contents of perceptual experiences are purely general then what kinds of entities do they represent? The simplest answer is that the content of a perceptual experience is an existential generalization representing the instantiation of various properties. For instance, consider the following three experiences:

---

<sup>59</sup> Other generalists include the following: Lewis (1980), McGinn (1982), Harman (1990), Millar (1991), Davies (1992), Siewert (1998), Byrne (2001), Hill (2009), Brogaard (2010, 2011), and Jackson (2012a, 2012b).

<sup>60</sup> See Bach (2010) for a detailed discussion of what it is for a content to be object-dependent.

- (1)  $S$  perceives  $a$  as having properties  $F$ ,  $G$ , &  $H$ .
- (2)  $S$  perceives  $b$  (where  $b \neq a$ ) as having properties  $F$ ,  $G$ , &  $H$ .
- (3)  $S$  undergoes a complete hallucination as of an object with properties  $F$ ,  $G$ , &  $H$ .

As far as our generalist is concerned, these are all cases in which the content of  $S$ 's experience is as follows:  $\langle (\exists x)(Fx, Gx, Hx) \rangle$ . What distinguishes (1), (2), and (3) from one another is thus a *content-independent* matter for the generalist. In particular, it is a matter of whether the subject perceives any objects, and if so, which ones she happens to perceive.

Let's turn now to the second question posed above: namely, can we determine whether a perceptual experience is accurate or not without determining which objects, if any, the subject perceives? Grice (1961) argues that there must be a causal dependence between an object and your visual experience in order for you to qualify as seeing that object. And even though many reject the claim that this causal dependence is sufficient for seeing an object, the claim that it is at least *necessary* is widely accepted.<sup>61</sup> For our purposes, the important point is that if you are a generalist then accepting a causal constraint on seeing drives a wedge between the question of whether the subject sees any objects, and the question of whether her experience is accurate. Why? Because it could happen that the subject is having a completely hallucinatory experience, despite the fact that the experience's content is satisfied by an object that the subject fails to perceive (due to the absence of the requisite causal relation). These kinds of experiences are referred to as *veridical hallucinations*.

---

<sup>61</sup> Proponents of a causal constraint on seeing include the following: Jackson (1977), Lewis (1980), McGinn (1982), Searle (1983), Millar (1991), Davies (1992), Tye (2009), Burge (2010), and Schellenberg (2010, 2011).

### 3.2.2 Particularism

In contrast to the generalist, the particularist accepts the thesis that whether a perceptual experience is accurate or not depends on whether the subject perceives any objects. However, only some particularists endorse the stronger thesis that if the subject perceives an object then that object enters into the content of her experience. Let's call those particularists who take on this additional thesis, *strong particularists*.

### 3.2.3 Strong particularism

Take the first experience described above, in which our subject perceives  $a$  as instantiating  $F$ ,  $G$ , and  $H$ . According to the simplest version of strong particularism, its content is the following Russellian proposition:  $\langle Fa, Ga, Ha \rangle$ . Given that in the second case, the subject is perceiving a distinct object,  $b$ , her experience is construed as possessing the following distinct content:  $\langle Fb, Gb, Hb \rangle$ .

What about the hallucinatory counterpart of these two experiences? If our subject fails to perceive any objects then the content of her experience cannot have any as constituents.<sup>62</sup> According to Tye's *singular-when-filled* theory, the content of the subject's experience contains gaps in the hallucinatory case, thereby rendering it inaccurate (2009, 548). Why inaccurate rather than neither true nor false? Tye (2009, 548) claims that construing the gappy content as inaccurate best fits the intuitive idea that in the hallucinatory case, the world diverges from the way it seems to the subject.<sup>63, 64</sup>

---

<sup>62</sup> I'm setting aside unusual cases in which the subject seems to hallucinate an object that she has perceived in the past (e.g. her mother). See Johnston (2004) for a discussion of these sorts of cases.

<sup>63</sup> Tye (2009) endorses the Russellian version of strong particularism. See Schellenberg (2010, 2011) for an alternative approach according to which the content of a perceptual experience contains both modes of presentation of properties and "object-related *de re* modes of presentation" (2010, 38). Importantly, these object-related *de re* modes of presentation are such that in

### 3.2.4 *Weak particularism*

Some proponents of particularism reject the strong variety and agree with the generalist that perceptual *content* is object-independent; however, unlike generalists, they are moved by arguments for the view that we cannot determine whether a perceptual experience is accurate or not without determining whether the subject perceives any objects.

For instance, on Burge's version of weak particularism, the content of a perceptual experience contains a demonstrative element (2010, 381). If there is no (appropriate) causal relation between this element and an environment particular, it fails to refer, and the subject thereby fails to perceive.<sup>65</sup> However, Burge denies that the content of a perceptual state contains a gap in those cases in which its demonstrative element fails to refer. Rather, the content of the hallucinatory experience is exactly the same as those of its successful counterparts. In each case, we have a demonstrative element, which happens to refer to an object in the successful case, and fails to refer to one in the hallucinatory case (2010, 381–3). Burge thus rejects the strong particularist's claim that those objects perceived by a subject are constitutive of her experience's content.<sup>66</sup>

The one thing that weak and strong particularists obviously agree on concerns the accuracy conditions of perceptual experiences, for according to weak particularists, if the

---

the non-hallucinatory case, the subject's experience has a token content that is constituted by the object(s) she perceives. In what follows, I set aside in-house disputes of this sort.

<sup>64</sup> In committing to strong particularism, Tye (2009) and Schellenberg (2010, 2011) are rejecting radical disjunctivism—in short, the thesis that perceptions and hallucinations do not share a common element. According to Tye (2009), perceptions and indistinguishable hallucinations represent the same properties; whereas, for Schellenberg (2010, 2011), they are constituted by the same modes of presentation of properties and objects. See Schellenberg (2010, 27–31) for a compelling critique of radical disjunctivism. McDowell (1982) appears to have been the first to articulate the thesis.

<sup>65</sup> As we saw in chapters 1 and 2, Burge also holds that a perceptual state must accurately represent certain features of an object—in particular, its cohesiveness, boundedness, and spatio-temporal continuity—in order to refer to it. Burge's demonstrative-like elements are thus akin to complex demonstratives, e.g. “That *F*.”

<sup>66</sup> See also Bach (2007) and Soteriou (2000) for defenses of weak particularism.

demonstrative element of a perceptual experience fails to refer then the experience counts as inaccurate (2010, 381–3). Weak particularists thereby join strong particularists in committing to the thesis that we cannot settle whether a perceptual experience is accurate or not without also settling which objects, if any, the subject happens to perceive.<sup>67</sup>

### *3.2.5 Adjudicating the debate*

How do we adjudicate the debate between generalists and particularists? There are four main ways in which theorists attempt to do so. One strategy is to argue that one side of the debate is best placed to capture the phenomenology of perceptual experience. A similar strategy is to argue that one side of the debate best captures the intuitive accuracy conditions of perceptual experiences. Other theorists take a different tack by focusing on the role that perceptual experiences play in grounding singular thoughts about environmental objects. Similarly, some theorists focus on the role that perceptual experiences play in grounding our knowledge of those same objects.

In what follows, I will address each strategy in turn, arguing that generalists and particularists are equally well placed to explain all of the relevant phenomena. Then, in sect. 6, I will offer a diagnosis of this stalemate, arguing that it is a systematic one. This will pave the way for my own view according to which there is just no fact of the matter as to whether the objects we perceive are among the contents of perception.

---

<sup>67</sup> Burge's view bears obvious similarities to Searle's (1983). Searle accepts the thesis that the content of a perceptual experience is not constituted by the object(s), if any, perceived by the subject. He also agrees that if the subject fails to perceive anything then her experience counts as inaccurate. However, he departs from Burge in denying that perceptual experiences contain demonstrative elements that refer to perceived objects (in successful cases). He is able to hold this combination of views because he posits a self-referential component in the content of a subject's experience: "When I see a flower, part of the content of the experience is that this experience is caused by the fact that there is a flower there" (1983, 123). Thus, whether the experience as of a flower is accurate or not depends on whether a flower is successfully perceived.

### 3.3 Explaining phenomenology

In theorizing about the nature of perception, one aspect that needs explaining is its phenomenology. Your experience of the coffee cup possesses a rich array of phenomenal features. For instance, there is something it's like to experience the whiteness of the cup, as well as its shape and motion. In explaining these phenomenal features is there anything to choose between generalism and particularism?

#### *3.3.1 The generalist's argument from phenomenology*

In making the case for generalism, McGinn provides the following argument from phenomenology:

... the content of experience is not to be specified by using terms that refer to the object of experience, on pain of denying that distinct objects can seem precisely the same: so when we are describing the content of an experience we should not make singular reference to the object of the experience ...<sup>68</sup> (1982, 51)

What McGinn is arguably right about is that we can explain the subjective indistinguishability of two experiences in terms of a purely general content. The problem with his argument, though, is that we can also explain it in terms of the type of content posited by the particularist.

In order to see why, consider the following subjectively indistinguishable experiences: in the first, you perceive *O* as a red sphere, one foot in front of you; in the second, you perceive numerically distinct object *P* as a red sphere, in the same location. So here we have two subjectively indistinguishable experiences, even though the subject perceives a different object in each case.

---

<sup>68</sup> See Davies (1992, 26) for a very similar argument.



What *is* held constant across the two experiences, and what does therefore explain their subjective indistinguishability, are the facts concerning which properties each one represents as being co-instantiated (as well as *where* each one represents them as being co-instantiated). More specifically, it is the fact that each experience represents the co-instantiation of redness and sphericity (at the same location relative to you, the subject) that explains why what it's like to undergo the first experience is the same as what it's like to undergo the second one. This is an important point: ordinarily, we do not undergo experiences as of uninstantiated properties (if ever). The fact that we do not is reinforced by pairs of subjectively distinguishable experiences, such as the following: (1) *Your experience as of a red sphere in front and a blue square to the left*, and (2) *Your experience as of a blue sphere in front and a red square to the left*. In order to explain why these experiences are distinguishable for you, we need to appeal to the facts concerning which properties each one represents as being co-instantiated (as well as where they are represented as being co-instantiated).

Thus if we are to explain both subjective distinguishability and subjective indistinguishability, we clearly need to posit perceptual contents that are rich enough to capture property co-instantiation. And the problem with McGinn's argument is that the type of content posited by the particularist succeeds in giving us this explanatory payoff.

Take the weak particularist's claim that your experience as of a red sphere has the following type of content:  $\langle \underline{\textit{That is a red sphere, one foot in front}} \rangle$ .<sup>69</sup> Compare this to the type of content posited by the generalist:  $\langle \underline{\textit{There is a red sphere, one foot in front}} \rangle$ . In both cases, redness and sphericity are represented as being co-instantiated (at a given location), and that is precisely what we need in order to explain what the

---

<sup>69</sup> In what follows, I will adopt the convention of having an English sentence in angled brackets denote the Russellian proposition expressed by that sentence.

experiences have in common with their subjectively indistinguishable counterparts, as well as what they lack in common with their subjectively distinguishable ones.

Notice that the explanation of subjective indistinguishability that I have just offered on behalf of the particularist does not commit her to object-dependent phenomenology, for it only appeals to those aspects of content that are shared between distinct tokens of the same type. Moreover, this point holds true regardless of whether the particularist construes these tokens along weak, or alternatively, strong particularist lines: either way, what we get is a type of content whose tokens all represent the co-instantiation of perceptible properties (at various locations).

Of course, some philosophers have argued that phenomenology is, in fact, object-dependent. In particular, some have claimed that our capacity to acquire demonstrative thoughts about environmental objects, as well as our capacity to acquire knowledge of those same objects, requires the existence of object-dependent phenomenology (e.g. see Campbell 2002). Schellenberg (2010, 2011) is a particularist who has recently provided arguments for the object-dependency of perceptual *content*—not phenomenology—by appealing to these very explananda. In sect. 5, I will argue that Schellenberg’s arguments fail to motivate object-dependent perceptual content, though, and that the generalist is equally well placed to explain the relevant phenomena. This will therefore double as an argument against the view that we must posit object-dependent *phenomenology* in order to explain these same phenomena.<sup>70</sup>

So there do not seem to be any compelling arguments from phenomenology in favor of either generalism or particularism. In order to settle the debate we will have to look elsewhere. In the next section, I want to shift the focus to intuitions about accuracy conditions.

---

<sup>70</sup> See Mehta (2014) for a recent argument against object-dependent phenomenology.

### 3.4 Intuition pumps and accuracy conditions

Do intuitions about the accuracy conditions of perceptual experiences favor one side of the generalism/particularism debate? More importantly, do intuitions of this sort have evidential value?

#### *3.4.1 The particularist's argument from intuitive accuracy conditions*

In marshalling support for the view that perceptual experiences have particularist accuracy conditions, Tye (2009, 544) asks us to consider the following scenario. You are looking at a mirror, which, unbeknownst to you, has a yellow cube behind it—perhaps you fail to realize that you are looking at a mirror in the first place. Of course, you do not see this particular cube, but there is also a white cube over your shoulder that is being reflected in the mirror. In fact, the lighting conditions are such that this second cube looks yellow to you. Moreover, it also looks to occupy the very position that the first one does (behind the mirror). There is an intuitive pull to say that your visual experience is inaccurate in this case—more strongly, that it misrepresents the location and color of the cube that you see (i.e. the one over your shoulder). And that is just to say that there is intuitive support for particularism.<sup>71</sup>

#### *3.4.2 Problems with the argument from intuitive accuracy conditions*

There are two main problems with this way of arguing for particularism. First, the generalist will no doubt reply by focusing on cases in which intuitions pull in the other direction. For instance, suppose you are hallucinating a banana. By sheer coincidence there happens to be one in front of you, matching the qualitative aspects of your experience right down to the smallest detail (e.g. color, shape, and orientation). Even though your experience is a complete hallucination, there is an intuitive pull to say that it is

---

<sup>71</sup> See Soteriou (2000) for a very similar case.

accurate nonetheless. After all, it “matches” the scene before your eyes perfectly—or so the generalist will insist.<sup>72</sup> Importantly, the generalist will not deny that your experience is defective in this case: it is defective, albeit accurate, because you are successfully perceiving neither the banana nor any other environmental objects. And as we saw above, the reason why the generalist can say this is because she is committed to a causal constraint on seeing.

Thus when it comes to intuitions about accuracy conditions, we seem to have a standoff between the generalist and the particularist. This will be important below, in sect. 6, where I argue that even if we were to construe the kinds of intuition pumps given above as having evidential value, they would actually provide evidence for the indeterminacy view: not generalism or particularism.

The second problem is that it is notoriously difficult to distinguish between the contents of one’s perceptual experiences, and the contents of those immediate perceptual *beliefs* they tend to cause. This casts serious doubts on the reliability of the intuitions invoked above. For instance, the intuition that I am misrepresenting the location of the cube over my shoulder could very well be engendered by the fact that I am disposed to token perceptual demonstrative *beliefs* about it—beliefs that misrepresent its location.<sup>73,74</sup> By the same token, the generalist-friendly intuition that I can have accurate hallucinatory experiences could well be engendered by the fact that I am disposed to acquire true existential beliefs in the relevant scenarios. For instance, in the case de-

---

<sup>72</sup> Tye (2009, 548–9) is a particularist who shares this intuition.

<sup>73</sup> I address the important question of how we acquire demonstrative beliefs about seen objects below, in sect. 5. Of course, the generalist will deny that we do so by undergoing perceptual experiences with demonstrative elements.

<sup>74</sup> The fact that intuitions are often unreliable when it comes to distinguishing between the contents of experience and the contents of immediate perceptual belief is further evidenced by the controversy over whether natural kind properties (e.g. *being a pine tree*) are among the contents of perception. See Siegel (2011) for a detailed defense of the positive view. See Prinz (2013) for a reply.

scribed above, I am disposed to acquire the true belief that there is a banana in front of me, even though I fail to perceive the banana that is, in fact, in front of me.<sup>75</sup>

### 3.4.3 Tye's argument from looks reports

So it appears unlikely that intuitions about accuracy conditions will help us to settle the debate between generalists and particularists. What about looks reports? Might they enable us to home in on the contents of perceptual experience? Tye certainly thinks so:

The strongest support for the view that perceptual experiences have singular contents into which the seen objects enter is the thought that when we see those objects they look some way to us, together with the further thought that an object can only look some way if it is experienced *as* being that way. (2009, 554)

The generalist can grant Tye's plausible claim that when you see an object it looks some way to you. What she will deny is that either the contents or accuracy conditions of visual experiences are fixed by those of true looks reports. And she can happily deny this because she is already committed to the view that none of the causal relations holding between a subject's token experience and the objects she sees are constitutive of either the experience's content or its accuracy conditions.

For instance, take a simple case in which *O* looks red to you. There is nothing stopping the generalist and the particularist from agreeing on what determines the fact that *O* looks this way to you. Plausibly, it involves your perceptual representation of *redness* standing in the requisite relation to *O* (i.e. the relation that constitutes your seeing of

---

<sup>75</sup> Tye (2009, 557) uses considerations of this very sort to explain away generalist-friendly intuitions about the accuracy conditions of perceptual experiences (which he shares). But I fail to see why analogous considerations would not also cast doubt on the reliability of those intuitions that are marshaled in support of particularism.

*O*). What the generalist and the particularist *must* disagree on is whether *O*'s looking red to you guarantees that the content of your experience is (partly) constituted by *O*. And how do we go about deciding that issue? By assessing the relevant explananda (i.e. those considered elsewhere in this chapter), and determining whether they require us to construe the objects that we stand in the looks relation to as partly constitutive of our experiences' contents. Without an independent argument for construing the looks relation as such, the particularist has no grounds on which to insist that the contents (and accuracy conditions) of visual experiences can be read off true looks reports.

Moreover, if we cannot settle the generalism/particularism debate by simply reading the contents of visual perception off true looks reports, the generalist has another compelling explanation of those cases that seem to provide intuitive support for particularism. Take Tye's scenario in which the white cube over your shoulder looks yellow. The generalist can easily accommodate the fact that the cube over your shoulder looks yellow—even though it is white—by appealing to the fact that *it* is the cube that you see: you don't see the one in front of you (behind the mirror) because the requisite causal relation is not instantiated. Tye claims that intuitions favor the view that your experience is inaccurate in this scenario; however, it is not clear why we should construe the intuition that is precipitated by this case as tracking the content of your experience, rather than just the content of the corresponding looks report, and if it is tracking the latter then the generalist can easily accommodate it. In other words, it seems plausible that the intuition pumped by Tye's example is that the cube over your shoulder looks to have a property that it does not in fact have, and *that* is an intuition that the generalist can readily accommodate.<sup>76</sup>

---

<sup>76</sup> Furthermore, Pautz (2009) argues—persuasively, I think—that we cannot read the contents of perception off true looks reports, because doing so would trivialize other important debates about the contents of perception: e.g. the debate over whether natural kind properties are

### 3.4.3.1 Tye's reply

Tye anticipates the kind of reply I just gave. More specifically, he considers a generalist who, in response to his argument from looks reports, defends the thesis that for  $O$  to look  $F$  is just “for  $O$  to cause (in the way involved in seeing  $O$ ) an experience of an  $F$ ” (2009, 558). Tye's response is to offer what he regards as a *reductio* of this view. He asks us to imagine a case in which he sees two objects,  $a$  and  $b$ , such that  $a$  looks red to him, and  $b$  looks green:

The object  $a$  causes my experience of something green, and it does so in the way involved in seeing  $a$ . But equally, so does  $b$ . In seeing both  $a$  and  $b$ , I undergo an experience that represents something green, on the existential proposal. This experience, however, also represents something red. So  $b$  not only causes my experience of something red but it also causes my experience of something green, and it does so in the way involved in seeing  $b$ . So  $b$  looks red, but it also looks green. Clearly something has gone wrong. (2009, 559)

According to Tye, what has gone wrong is that the generalist has “removed the seen object from the content of the experience involved in seeing it” (2009, 559).

But removing the seen object from the experience's content is clearly not what has gone wrong in Tye's example. On the contrary, what has gone wrong is the account of looks reports that he has suggested on the generalist's behalf. As was mentioned above, according to Tye's proposal, the generalist can adopt the view that for  $O$  to look  $F$  is for  $O$  to cause (in the way involved in seeing) *an experience* of an  $F$  (2009, 558). But why would a generalist want to accept this account of looks reports? In adequately capturing

---

represented in perception; the debate over whether perceptual states are representational in the first place; and so on.

the fact that *b* looks green, and not red, the generalists will want to appeal to the more fine-grained causal relation that holds between *b* and *that part of her experience* for which *b* is causally responsible (namely, the part that represents greenness, not redness). That is to say, the generalist should adopt the view that for *O* to look *F* is for *O* to cause *part* of the subject's experience to represent an *F*. Adopting this approach enables her to adequately capture the fact that it is *O* that looks *F* (not some other object that is partly causally responsible for the experience, and that does not look *F*). It also enables her to adequately capture the fact that *O* looks *F* (and does not look to have some other property that a distinct object causes her experience to represent).

In a footnote, Tye very briefly registers the fact that “further conditions could be imposed” to deal with his attempted *reductio*. He doubts that these conditions would help the generalist, though:

... these conditions would require there to be a sensitivity in the relevant part of the content of the token experience to the relevant object. Given such a sensitivity, it is no longer clear that the object has not been smuggled into the content after all. (2009, 559)

Tye's remarks are dialectically ineffective, though, for the generalist claims to have independent reasons for denying that *any* of the causal relations holding between a subject's experience and objects in her environment are content-conferring ones: this is precisely what the appeal to the possibility of veridical hallucination is designed to show. It is for this reason that the account of looks reports suggested on behalf of the generalist—in terms of the particularly *fine-grained* causal relations that hold between certain parts of the subject's experience and certain environmental objects—is not “smuggling” those objects into the content of the experience. If the possibility of veridical hallucina-



tion provides a reason to deny a content-determining role for *any* of the causal relations that hold between a token experience and objects in the subject's environment, then it obviously provides a reason to deny this role for coarse- and fine-grained ones alike. Tye's worry is thus a red herring.

For all these reason, I am highly doubtful that intuition pumps and looks reports can aid us in settling the debate between generalists and particularists. Thus, in the remainder of this chapter, the focus will be on the explanatory roles that perceptual experiences are widely seen as playing with respect to our cognitive and epistemological lives: roles that could provide us with the constraints on perceptual content that we are after.

### **3.5 Demonstrative thoughts and knowledge of particulars**

One argument that has been proposed in support of particularism concerns our capacity to acquire demonstrative thoughts about the objects we perceive. For instance, Martin (2002, 182) states that generalism is "at odds with the simple thought that our experiences ground demonstrative judgements about particular objects." He goes on to add,

As Moore was keen to stress, a glance at one's desk may lead to the judgement, '*That* (directing one's attention at one thing on one's desktop) is an inkstand'. One is not stuck with merely the possibility of judging that there is some inkstand in the vicinity, one can pick out the very inkstand in question and make a demonstrative judgement about it.<sup>77</sup> (2002, 182)

But why exactly can't we explain the capacity to acquire demonstrative thoughts about

---

<sup>77</sup> See also Campbell (2002), Brewer (2006), and Johnston (2006).

perceived objects if we construe perceptual experiences as having (object-independent) existentially quantified contents? In what follows, I will argue that the generalist is equally well placed to explain this capacity.

### 3.5.1 *The argument from the one-many*

The first argument is premised on the claim that the perceiving relation is plausibly one-many, as opposed to one-one. For instance, in seeing the banana you also see its facing surface. So what determines the fact that you are seeing both of these objects at the very same time? If you are a particularist, and you think that the objects we perceive are demonstratively referred to by the experiences we undergo when we perceive them, there are two options. The first is to claim that the banana and its facing surface are referred to by distinct elements of your experience; on this view, the perceptual reference relation is one-one. The second option is to claim that these distinct objects are referred to by the very same element of your experience; on this view, the perceptual reference relation is one-many (i.e. it admits cases of plural reference). Let's assess each option in turn.<sup>78</sup>

The main worry with the first option is that none of the potentially reference-grounding relations that hold between the banana and a given element of your experience appear to be ones that do not also hold between *that very same element* and the banana's facing surface. For instance, suppose we were to construe the perceptual reference relation as a causal one, such that an element, *X*, of your token experience refers to an object just in case that object stands in the appropriate causal relation to *X*. But if the banana and its facing surface are both appropriate causes of parts of your experience, then it is very hard to see how they could fail to count as appropriate causes of

---

<sup>78</sup> I'm setting aside contextualism about object-perception for the moment. I explain how going contextualist affects the arguments I give in this section, below in sect. 3.5.6.

*the very same parts*. It is not as if the light that is reflected from the banana's facing surface is received at a different part of the retina than the light that is received from the banana itself. Thus if we go with the view that the perceptual reference relation is a causal one, it is very hard to see how an element of your experience could refer to an ordinary object without also referring to its facing surface.

Arguably, a more promising approach is to construe the perceiving relation as a one-many relation. For instance, if we construe it as a causal one then it is plausible that the element of your perceptual experience that puts you into the seeing relation with the banana, also puts you into the seeing relation with its facing surface. Moreover, there are good reasons for construing the seeing relation as a one-many relation even if we deny that it is a purely causal one, for it is widely held that to see the facing surface of an ordinary object just is to see the object itself. Thus, as long as an element of your experience puts you into the seeing relation with the banana's facing surface, this same element thereby puts you into the seeing relation with the banana itself.<sup>79</sup>

We are now in a position to see why the fact that the perceiving relation is likely one-many, as opposed to one-one, makes the generalist and the particularist equally well placed to explain our capacity to token thoughts that demonstratively refer to the objects we perceive. Suppose we want to explain the fact that, having seen the banana, you were able to token the thought *THAT IS A BANANA*. If the particularist concedes that the element of your perceptual experience that picks out the banana also picks out its facing surface, then in order to explain how you acquired a thought that determinately picks out the banana, and not its facing surface, some sort of reference-fixing

---

<sup>79</sup> Proponents of the thesis that we see 3D objects by seeing their facing portions include the following: Jackson (1977, 19), McGinn (1982, 54), McLaughlin (1984), Bermudez (2000), and Kriegel (2009, 225).

mechanism needs to be posited: one that weeds out the unwanted referent (i.e. the banana's facing surface).

For instance, perhaps your thought determinately picks out the banana, but not its facing surface, because it contains the complex demonstrative 'That *B*-shaped object,' which contains a predicate ('is *B*-shaped') that is satisfied by the banana, but not its facing surface.<sup>80</sup> This is not the place to weigh up specific accounts of what fixes the reference of perceptual demonstrative thoughts. The important point is that if the seeing relation is one-many then the particularist cannot claim that demonstratively referring thoughts simply inherit their referents from perceptual experiences via a straightforward causal dependence: some sort of semantic mechanism needs to work on the plurally referring experience and home in on the relevant referent so as to produce a *singularly* referring thought, and whatever mechanism that may be, it is one that the generalist can help herself to as well. At the very least, the onus is on the particularist to show otherwise.<sup>81</sup>

### 3.5.2 *The two-vehicle alternative*

Admittedly, I have not given a knockdown argument for view that the seeing relation is one-many, and there are certainly accounts of seeing that are potentially at odds with it. According to one such account, the seeing relation is a causal-cum-satisfactional one. More specifically, the view is that *S* sees *O* just in case (i) *O* is an appropriate cause of *S*'s visual state, *V*, and (ii) *V* is sufficiently accurate of *O*.<sup>82</sup> If this account of seeing is correct then it may well be that the perceptual vehicle that puts the subject into the

---

<sup>80</sup> See Strawson (1959) and Dummett (1973, 1981) for versions of this view.

<sup>81</sup> In locating a semantic mechanism that does the job of homing in on just one of the particulars that is perceived by the subject (e.g. the banana, as opposed to its facing surface), the descriptivist approach is just one potential solution among many. For an account that is neutral between generalism and particularism, and that is also non-descriptivist, see Dickie (2011).

<sup>82</sup> See Burge (2010) for a version of this view.

seeing relation with the banana is distinct from the one that puts her into the seeing relation with its facing surface. Simplifying things a little, one vehicle might carry a content of the form, *<That B-shaped body>*, while the other carries a content of the form, *<That F-shaped surface>*, where *being B-shaped* is a property instantiated by the banana, and *being F-shaped* is a property instantiated by its facing surface, but not vice-versa. Call this *the two-vehicle approach*.<sup>83, 84</sup>

If the two-vehicle approach is workable, it paves the way for a view according to which demonstrative thoughts inherit their referents from the experiences that precede them in virtue of a straightforward causal dependence. More carefully, call the vehicle that perceptually refers to the banana (not its facing surface), *B*. And call the distinct vehicle that perceptually refers to the banana's facing surface (not the banana), *S*. Given that *B* uniquely refers to the banana, this opens up the possibility of a view according to which the demonstrative element of your *thought* refers to the banana, and to the banana alone, because it was appropriately caused by *B*. That is to say, there would be no need to posit a semantic mechanism that weeds out the unwanted referent (the banana's facing surface) and homes in on the banana: that job has already been done by the vehicle that carries the 3D-banana-shape representation. Call this *the inheritance view*.

---

<sup>83</sup> One way of thinking about these vehicles is to construe them as occurring at different stages of visual processing: the vehicle that represents the property of *being an F-shaped surface* occurring at "the surface-based stage," with vehicle that represents the property of *being a B-shaped body* occurring at the "object-based stage." See Frisby & Stone (2010, 155–204) for a discussion of relevant issues.

<sup>84</sup> The two-vehicle approach is not without its problems, for what is needed is an account of how the visual system produces representations as of 3D objects and distinct representations as of their facing surfaces. Needless to say, providing such an account is not an easy task, and the moral of chapter 1 was that it is a futile one: visual states are indeterminate as between 3D objects and their facing surfaces.

In responding to the proponent of the inheritance view, there are two things that a generalist can say. First, if the proponent of the inheritance view can posit complex-demonstratives in perception, why can't the generalist make the same kind of move at the level of perceptual demonstrative *thought*? In other words, why can't the generalist tell a reference-fixing story for perceptual demonstrative thoughts by construing *them* as carrying demonstratives with attributive elements: in particular, attributive elements that have a reference-fixing role to play?

In any case, it is not clear why the generalist cannot just give the same reference-fixing story (for perceptual demonstrative thoughts) as the proponent of the inheritance view, minus the claim that the perceiving relation is a content-conferring one (with respect to perceptual experience). The generalist is certainly free to adopt a causal-cum-satisfactional theory of the seeing relation. What she will deny is that standing in this relation to an object amounts to perceptually representing it. Thus, having secured a perceptual vehicle—namely, *B*—that puts the subject into the seeing relation with the banana, and the banana alone, the generalist could follow the proponent of the inheritance view by telling a reference-fixing story for the subject's demonstrative thought by appealing to the very same causal dependence between *B* and the demonstrative element of that thought.

### *3.5.3 Contextualism (again)*

In chapter 2, I argued that whether we see objects, in addition to their facing surfaces, is a context-sensitive affair. So far, in responding to the argument from demonstrative reference, I have not been assuming contextualism. So how does the argument from demonstrative reference fare if we bring contextualism back into the picture?

In fact, I effectively addressed this question in chapter 2. There, I argued that the contextualist can provide an explanation of how it is that a subject, *S*, acquires demon-

strative thoughts with context-invariant referents, by appealing to context-invariant facts about *S* and her environment. In particular, I pointed out that the contextualist is free to help herself to the kinds of resources described above. For instance, she could provide an account that appeals to complex demonstratives (in thought). The key point is that if we combine contextualism with particularism, what we get is a view according to which the referents of visual states vary according to the interests of ascribers. Thus, if we are to tell a story about how perceptual demonstrative thoughts acquire context-*invariant* referents, we must posit a mechanism that homes in on the relevant referent; namely, a mechanism that appeals to context-invariant facts about the subject and her environment. And whatever those factors are, they are there for generalists to help themselves to.

#### 3.5.4 *The strong particularist's reply*

I have just argued that the particularist does not enjoy an advantage over the generalist when it comes to explaining the capacity to think, demonstratively, about seen objects. More specifically, I have argued that if the particularist provides an explanation of this capacity, the generalist can simply proffer the same explanation, minus the claim that the object seen is part of the given experience's content. At this point, the *strong* particularist might reply as follows: "But surely, if we construe the seen object, *O*, as *constitutive* of the given experience, we have a clear advantage when it comes to explaining how the demonstrative element of the resulting thought, *T*, acquired *O* as its referent: after all, *T* was caused by an experience that is constituted by *O* itself!"

If what I have argued is right, the strong particularist's insistence lacks dialectical force. What we need is an independent argument for the thesis that appealing to the perceiving relation does not suffice to explain how *T*'s demonstrative element acquired *O* as its referent, *unless the perceiving relation is being construed as constitutive of the*

*experience that caused T*, but no such argument has been forthcoming. The metaphysical question of whether perceptions are constituted by their objects seems orthogonal to the issue of how perceiving an object can engender demonstrative thoughts about it.

In order to illustrate this point, consider the following analogy. We want to explain why Joyce has green eyes, and so we gather the requisite information about the color of her parents' eyes, along with other relevant facts about their genes and physiology. Now, suppose we are having an argument on the side about whether a person could have had biological parents other than the ones she in fact has: you think that a person's biological parents are essential to them, whereas, I do not. Does this disagreement about whether Joyce could have had different parents have any bearing on our explanation of how it is that she came to have green eyes? Presumably, it does not. We agree that Joyce's eye color can be explained solely in terms of the fact that she was born of certain parents, with certain genetic constitutions, and so on. The issue of whether she could have existed with different parents is just irrelevant: it is a metaphysical side issue that must be decided on independent theoretical grounds.

The same is true of the way in which we explain Joyce's capacity to token thoughts that demonstratively refer to the objects she perceives. The full explanation of how Joyce came to token a demonstrative thought about *O* may well include an appeal to the fact that she perceived *O*. However, in giving this explanation, we don't need to have settled the issue of whether Joyce could have had *that very same experience*—the one that actually put her into the perceiving relation with *O*—while being perceptually related to a different object. Like the issue of whether she could have been related to different parents, this is an issue that must be decided on independent theoretical grounds.



### *3.5.5 The epistemological argument*

I have just argued that the particularist does not have an advantage when it comes to explaining our capacity to acquire demonstrative thoughts about seen objects. But if that is the case, it becomes very hard to see how the epistemological argument even gets off the ground.

Take Schellenberg's (2011, 735–9) version of the argument.<sup>85</sup> She begins with a simple example in which Anna sees a coffee cup and thereby gains perceptual knowledge of it. Plausibly, if she had perceived a distinct, albeit indistinguishable, cup she would have gained knowledge of that cup instead. According to Schellenberg, the generalist cannot explain Anne's capacity to acquire these distinct pieces of knowledge:

If her experience would be the same whether or not she were perceiving the particular coffee cup that she is in fact perceiving, it is not clear how her experience could ground knowledge of that particular coffee cup ... (2011, 735)

She goes on to add:

... experience can only ground knowledge of particular objects, if the particular object to which the subject is perceptually related makes a constitutive difference to the experience. For only if experiences are individuated by their objects, can perceptual experiences be the evidential basis of knowledge of particular objects. (2011, 735)

According to Schellenberg, it follows that the strong particularist has an advantage over her competitors when it comes explaining the capacity to gain perceptual knowledge of environmental objects.

---

<sup>85</sup> See also Campbell (2002, ch 2).

However, in light of the reasons, adduced above, for thinking that generalists and particularists are equally well placed to explain the capacity to acquire thoughts that demonstratively refer to perceived objects, the generalist has a ready reply. Take the simple case in which Anne sees the cup and acquires the knowledge that *that cup is white*. How will the generalist explain Anne's epistemological feat?

The generalist can start by explaining how perceiving the cup determined the fact that Anne's subsequent thought demonstratively referred to it. Next, she can appeal to her preferred account of whatever additional ingredients must be in place for this demonstrative thought to count as a piece of knowledge. For instance, she might appeal to the fact that Anne has no reason to believe that she is undergoing a visual illusion; she might appeal to the reliability of Anne's thought; and so on. The key point is that her experience *does* provide evidence for her belief that *that cup is white*. It is just that other ingredients—over and above her experience alone—must be in place for this belief to get elevated to the level of knowledge. It is not clear what, according to Schellenberg, is missing from this explanation. In order to secure an adequate evidentiary role for Anne's experience, why must we commit to the additional claim that the cup is constitutive of it?

At one point, Schellenberg imagines a generalist replying to her argument by insisting that Anne's knowledge of the cup is grounded in the causal relation that holds between it and her experience. Schellenberg raises the following worry about this approach:

The problems with this causal strategy are the same as the ones that face any causal view of experience in general and any causal view of knowledge in particular ... it is far from clear that causal relations can be analyzed such that they play an epistemological role. (2011, 737)

But why can't the generalist just agree that causal theories of both experience and knowledge are to be rejected? For instance, if the generalist is moved by standard worries about deviant causes, why can't she adopt the view that the seeing relation is a non-causal primitive one? The generalist could then construe this primitive seeing relation as grounding Anne's singular thoughts about the objects she sees. It is just that it would not be construed as having any role to play in fixing the content of her experience.

The question would then arise as to which of Anne's singular thoughts count as knowledgeable ones, and if causal theories of knowledge are to be rejected then our generalist is free to adopt an alternative account of the ingredients that must be in place: she is certainly not wedded to a causal theory. For instance, there is no reason why our generalist could not adopt a virtue-theoretic approach to knowledge; a non-causal version of the JTB theory; and so on. The onus is on Schellenberg to show that if we adopt anything other than a causal theory of knowledge, then Anne's thought about the cup can only be construed as a knowledgeable one if her *experience* is constituted by the cup, and nothing Schellenberg says goes any way towards establishing this claim.<sup>86</sup>

### 3.6 There is no fact of the matter

In examining those phenomena that promised to settle the debate between generalists and particularists, a stalemate has emerged. In particular, a stalemate has emerged on

---

<sup>86</sup> Schellenberg (2011, 735) also mentions, approvingly, Campbell's (2002, ch. 2) claim that perceptual experience grounds our knowledge of which particular objects our demonstratives refer to, and that in order to explain this capacity we must construe experiences as being individuated by their objects. Again, no argument been given for thinking that the subject's experience must be *constituted* by the object perceived in order for this capacity to be explained. It is the fact that the subject's experience is perceptually related to a specific object—namely, the one demonstrated—that is doing the explaining. Whether that perceiving relation has an individuating role to play is something that must be decided on independent theoretical grounds.

two main fronts.

The first front concerns phenomenology. Generalists have traditionally been regarded as having the upper hand when it comes to explaining this aspect of perceptual experience; however, as I have argued, generalists and particularists are actually on an equal footing here.

The second front concerns the roles that perceptual experiences play in grounding demonstrative thoughts about seen objects, as well as knowledge of those same objects. Particularists have traditionally been regarded as having the advantage when it comes to accounting for these roles. However, as I argued above, whenever the particularist explains some cognitive or epistemological capacity by appealing to an object-dependent content, the generalist can adopt the same explanation, minus the claim that the perceiving relation is a content-conferring one (with respect to perceptual experience).

This last point is reinforced by the fact that generalists and particularists can agree on what constitutes the seeing relation: perhaps it is a purely causal relation; perhaps it is a causal-cum-satisfactional relation; or perhaps it is a primitive one. The key point is that the perceiving relation is what puts us into cognitively rewarding contact with environmental objects, and it does so regardless of whether we construe it as a demonstrative reference-relation, or, as a non-demonstrative relation that is coupled with an existentially quantified content.

Moreover, it is important to realize that the kind of stalemate that we get is not one in which, on the balance of things, the evidence is neutral between generalism and particularism because the unique advantages enjoyed by one side are cancelled out by those enjoyed on the other. If that were the case, the most plausible conclusion to draw would be that more evidence is required. But that is not the kind of stalemate that we have reached. The kind of stalemate that we have reached is one in which both sides can explain phenomenology, and both sides can also explain how the perceiving relation puts

us into cognitively rewarding contact with environmental objects. It therefore appears unlikely that further evidence will break the deadlock.

As an example, consider the claim that studies of multiple-object tracking provide evidence for particularism. In particular, consider Pylyshyn's (2001, 2003, 2006, 2007) claim—outlined in chapters 1 and 2—that multiple-object tracking is realized by a mechanism that picks out targets and maintains referential contact with them via the assignment of “visual indexes”: singular elements akin to pure indexicals. A detailed assessment of whether the data support Pylyshyn's thesis is beyond the scope of this dissertation. Nonetheless, I want to make two brief points.

First, I see no reason why the systematic stalemate between the generalist and the particularist will not extend to the explanation of multiple-object tracking. In accommodating the data yielded from relevant studies, the generalist will just claim that the causal-tracking relations that obtain between a subject's successive visual states and target objects are not content-conferring ones; rather, they are just successive instantiations of the seeing relation. Second, even if one rejects this suggestion, the point remains that studies of multiple-object tracking do not allow us to adjudicate between strong and weak versions of particularism.

For these reasons, I think we should embrace the conclusion that there is just no fact of the matter as to whether our experiences represent the objects we perceive. What generalists and particularists are providing us with are equally acceptable ways of assigning contents to our perceptual states.

### *3.6.1 Intuition pumps and salient content assignments*

In order to illustrate the indeterminacy view I am defending, reconsider the kinds of examples that generalists and particularists typically use to pump intuitions about accuracy conditions. In the case of apparently veridical hallucinations, there is a temptation

to construe the subject's experience as accurate, despite the fact that she does not perceive any objects. On the other hand, there are cases in which the scene before the subject "matches" her experience, and yet she clearly misperceives some object (e.g. the cube that is reflected in the mirror in Tye's example). In these latter kinds of cases, there is a temptation to say that the subject's experience is inaccurate.

So we have these different intuitions that seem to favor distinct ways of assigning contents and accuracy conditions to our perceptual experiences. However, rather than construing them as incompatible, I am arguing that we should construe them as *compatible* intuitions about distinct, albeit equally acceptable, ways of assigning contents and accuracy conditions to our perceptual experiences.

In particular, we should see the standard thought experiments that generalists and particularists appeal to as making different, albeit equally acceptable, content assignments *salient*. When we are asked to imagine apparent cases of veridical hallucination, it has been made salient to us that there are objects positioned exactly where the subject's experience says there are, instantiating exactly those properties that the subject's experience says they instantiate, and so in this way, the generalist's assignment of content becomes salient.

On the other hand, when we imagine cases in which the subject clearly misperceives some object in her environment, the particularist's assignment of content becomes salient. This is reinforced by the fact that we are accustomed to characterizing misperceptions in terms of object-dependent looks (and sees-as) reports. For instance, we say things like: "That ball looks red to Roger, but it's actually pink"; "Margaret is seeing Roger as located in front of her, but he's actually off to her left"; and so on. When we are confronted with cases of visual illusion, it is thus natural to think of them in terms

of those object-dependent content assignments that are made salient by the relevant looks and sees-as reports.<sup>87</sup>

### 3.6.2 *The irrelevance of parsimony*

Could it be that generalism is to be favored on grounds of parsimony? Take two simple cases. In the first, you see  $b$  as  $F$ . In the second, you undergo a hallucination as of an  $F$ . The generalist will assign the following content to both of your experiences:  $\langle (\exists x)(Fx) \rangle$ . On the other hand, the strong particularist will assign the token content,  $\langle Fb \rangle$ , to the first experience, and the token content,  $\langle F\_ \rangle$ , to the second. Is the generalist's content assignment more parsimonious than the strong particularist's (I address the weak particularist's below)? More importantly, if it is more parsimonious, does this have evidential value?

It is controversial whether parsimony has evidential value in the domain of philosophical theorizing, let alone in the hard sciences.<sup>88</sup> Elsewhere, I have argued that appeals to parsimony in semantic theorizing are especially problematic (Phillips 2012). However, we needn't introduce general worries about the role of parsimony in semantic theorizing in order to raise serious doubts about its suggested role in adjudicating the dispute between generalists and strong particularists.

First and foremost, notice that the generalist and the strong particularist can agree on the following:

---

<sup>87</sup> As I argued above, what we cannot do is infer from the fact that there are true object-dependent looks (and sees-as) reports, that perceptual content is object-dependent. To do so would beg the question against the generalist.

<sup>88</sup> See Huemer (2008) for an argument that ontological parsimony has no evidential value in what he calls "typical philosophical contexts."

- (1) The neurobiological complexity of the perceptual vehicles that carry the contents of our perceptual experiences.
- (2) The causal relations that hold between perceptual vehicles and environmental entities.
- (3) What constitutes the perceiving relation.

What the generalist and the strong particularist *must* disagree on is whether the perceiving relation is a content-conferring one (with respect to perceptual experience). But in insisting that we construe the perceiving relation as content-conferring with respect to perceptual experience, the strong particularist is not thereby insisting that we admit an additional head-world relation into our ontology. Rather, she is insisting that we take a relation that both parties to the debate already countenance, and construe it in a certain way: namely, as a content-conferring one. Once we recognize this point, it becomes very hard to see how an appeal to parsimony could give us any reason to favor generalism over strong particularism (let alone weak particularism).

If the strong particularist were committing herself to more complicated perceptual vehicles than the generalist, or, if she were positing causal relations (between perceptual states and environmental entities) that the generalist does not, then perhaps the generalist's appeal to parsimony would be compelling. But as we have just seen, the strong particularist takes on no such additional commitments.

Moreover, even if one remains convinced that generalism is to be favored over strong particularism on grounds of parsimony, it is important to keep in mind that the weak particularist provides a uniform assignment of contents to perceptions and their indistinguishable counterparts. For instance, in each of the cases given above, the weak particularist assigns a token content of the form,  $\langle \textit{That is an } F \rangle$ , while the generalist posits a token content of the form,  $\langle \textit{There is an } F \rangle$ . Thus even if considerations of parsi-



mony provided a reason to favor generalism over strong particularism, they would not provide a reason to favor generalism over weak particularism. In other words, the indeterminacy view would remain the best choice: it is just that we would be restricting the equally acceptable content-assignments to those offered up by generalists and *weak* particularists.

Finally, suppose we construe intuitions about the accuracy conditions of perceptual experiences as having evidential value. In that case, there is another way of replying to the worry about parsimony. For even if the generalist can lay claim to providing the most parsimonious assignment of contents to perceptions and their subjectively indistinguishable counterparts, that would only constitute a reason to favor generalism if there were not a third account that is more explanatorily powerful than *both* generalism and particularism. But if we take intuitions about accuracy conditions as having evidential value, then the indeterminacy view is just such an account, for it provides us with the most straightforward explanation of the fact that intuitions are pulled in the direction of both object-dependent and object-independent content assignments. Our intuitions are pulled in both directions because the content assignments in question are all equally acceptable—intuition pumps merely altering which assignment is the most salient. Given that the indeterminacy view provides the best explanation of our intuitions, considerations of parsimony arguably do not get a foothold—or so the argument goes.<sup>89, 90</sup>

---

<sup>89</sup> Above, I suggested that we should not construe the relevant intuitions as having evidential value. However, some theorists will no doubt disagree with me on this issue. My point here is just that if we *do* construe the relevant intuitions as having evidential value, they will provide support for the indeterminacy view.

<sup>90</sup> Perhaps a case can be made for the claim that the standard versions of generalism and particularism I have been focusing on are to be favored over Searle's (1983) account on grounds of parsimony, for recall that Searle goes beyond standard approaches and posits a self-referential component in the content of every perceptual experience. Addressing this issue in adequate detail is beyond the scope of this dissertation: although, see Phillips (2014) for a discussion of problems surrounding the notion of indirect representation.

### 3.6.3 *The multiple contents view*

Another worry for the indeterminacy view concerns the possibility that perceptual states possess multiple layers of content. In particular, why not adopt the view that perceptual experiences have two layers of determinate content—an object-dependent layer and an object-independent layer—rather than a single layer of indeterminate content? Call this *the multiple-contents view* (MCV). For instance, consider the version of MCV according to which your successful perception of  $b$  as  $F$  has two layers with the following contents:  $\langle (\exists x)(Fx) \rangle$  and  $\langle Fb \rangle$ . In the hallucinatory case, the content of your experience would have the following two contents:  $\langle (\exists x)(Fx) \rangle$  and  $\langle F\_ \rangle$ .<sup>91</sup> Are there any compelling reasons to favor MCV over my alternative according to which your experience possesses a single layer of indeterminate content? I don't see how there could be.

The extra layer of content that we get with MCV does not give us any explanatory power that we do not already have with a single layer of content. For as I have argued, the generalist and the particularist both posit content-types that suffice to explain subjective indistinguishability (as well as subjective distinguishability). Moreover, when it comes to the capacity for demonstratively referring thoughts, as well as the capacity to acquire perceptual knowledge of particulars, I have argued that we should adopt explanations that are neutral between generalism and particularism. Thus when it comes to all of these phenomena, positing an extra layer of content will not give us any explanatory payoff that we do not already have with a single layer. So that just leaves the issue of accuracy conditions.

Is it that we need two layers of content to explain our intuitions about the accuracy conditions of perceptual experiences? Take the issue of veridical hallucination. You are having a complete hallucination as of a blue sphere in front, and there happens to be one in the relevant position. MCV says that the object-independent layer of your expe-

---

<sup>91</sup> Chalmers (2006) defends a version of the multiple-contents view.

rience is satisfied, but that the object-dependent layer contains a gap and is therefore inaccurate. Do our intuitions favor the view that your experience is *partly* accurate and *partly* inaccurate in this way? Even if we grant the controversial claim that intuitions have evidential value here—and I have suggested that they do not—I doubt that they are fine-grained enough to motivate MCV. I certainly harbor the intuition that your experience is accurate “in one way” and inaccurate “in another way.” That is to say, I am pulled in the direction of the purely general content assignment, and I am also pulled in the direction of the object-dependent one. But that underspecifies whether these different “ways” correspond to different layers of content, or alternatively, equally acceptable content assignments.

Finally, even if the proponent of MCV managed to find a case in which, intuitively, the subject’s experience determinately possesses *both* accurate and inaccurate contents, at best that would provide us with a reason to augment our list of equally acceptable content assignments. That is to say, we would now have a reason to think that the content assignment provided by MCV is one of those equally acceptable assignments that can be made more or less salient depending on the intuition pump deployed.<sup>92</sup>

### 3.7 Context-sensitivity vs. indeterminacy

I have been arguing that it is indeterminate whether perceptual content is object-dependent: why indeterminate, and not context-dependent? More specifically, as was the case with object-seeing, why not say that the notion of perceptual content is con-

---

<sup>92</sup> Notice that, in the same way, claiming that “rabbit” is not referentially indeterminate because it determinately refers to *both* the set of rabbits and the set of undetached-rabbit-parts would not go any way towards undermining Quine’s (1960) inscrutability thesis. At best, it would just provide us with a reason to augment the list of equally acceptable extensions for the term “rabbit.” Thanks to David Rosenthal for some very helpful discussions of this issue.

text-sensitive: relative to some contexts, *S*'s state qualifies as having an object-dependent content; relative to others, it qualifies as having a purely general content.

As was mentioned above in chapter 2, Bence Nanay (2015) defends a variant of this view; however, he is concerned with the more general debate as to whether perceptions are, fundamentally, relations to objects in the world as opposed to object-independent representational states; whereas, I'm concerned with the in-house debate between those who agree that perceptual states are representational, but disagree as to whether their contents are object-dependent. Nonetheless, the worries I have about giving a contextualist resolution to the generalism/particularism debate carry over to the more general debate that Nanay is concerned with.

As I see it, the problem with going contextualist—as opposed to indeterminist—about the issue of whether perceptual content is object-dependent is that doing so would require us to uncover explananda that justify distinct content-assignments. That is to say, we would need to uncover at least one phenomenon, the explanation of which requires us to posit an object-dependent content; and at least one phenomenon, the explanation of which requires us to posit a purely general content. And that is just not what we have found. Rather, we have found that generalists and particularists are equally well placed to explain each of the relevant phenomena (e.g. phenomenology, the capacity to acquire perceptual demonstrative thoughts, and so on).

Contrast this situation to the case of object-seeing, which does, indeed, require us to assign different extensions relative to different explanatory contexts: for instance, recall that explaining the capacity for amodal surface completion requires us to invoke a head-world relation that is different from the one that explains the capacity for facial recognition: the extension of the notion of object-seeing thus varies with the explanatory context of the theorist, rather than simply being indeterminate. Given that the contents

offered up by generalists and particularists do not exhibit this kind of pattern, there is no justification for going contextualist, rather than just positing indeterminacy.

### **3.8 Conclusion**

In examining key phenomena that we might expect a theory of perceptual content to explain, it has emerged that there is a systematic stalemate between the different content assignments provided by generalists and particularists: whatever the particularist explains in terms of an object-dependent content, the generalist can explain (equally as well) in terms of a purely general content that is paired with a non-content-conferring perceiving relation. I have argued that we should therefore abandon the widespread assumption that perceptual content is either determinately general or determinately object-dependent. Instead, we should embrace the surprising conclusion that there is just no fact of the matter. But why does this constitute such a surprising conclusion? Let me conclude by offering some brief diagnoses.

One reason that the kind of content indeterminacy I am positing may have gone unnoticed is that it does not fall neatly into the categories of content indeterminacy with which we are most familiar. For instance, it is unlike Quinean inscrutability in several ways: most obviously, the indeterminacy that I have argued for does not involve fixing the truth values of sentences in a given language, and then varying the extensions of sub-sentential constituents in such a way that these truth values are preserved (Quine 1960). The indeterminacy I am positing is also not the kind that many see vague terms as exhibiting. Nor is it the kind arguably exhibited by theoretical terms that have survived significant changes in the theories in which they are embedded (see Field 1973).

Another possible reason why the kind of indeterminacy I am positing has been passed over is that many are likely assuming that if generalism and particularism both give us the same explanatory payoffs, then generalism is to be preferred on grounds of parsimony.

mony. But as I have argued, it is very doubtful that parsimony can play this role in adjudicating the dispute, and even it could, it would only allow us to rule out strong particularism: we would still be left with the equally acceptable content-assignments offered up by generalists and weak particularists.

This last point has implications for disputes about content well beyond the one I have engaged with in this dissertation. In weighing up apparently competing assignments of content to a given mental state or linguistic expression, we should be wary of appeals to intuition that are not bolstered by compelling accounts of why one of the assignments is more explanatorily powerful than its competitors. If a compelling account of this kind is not forthcoming, then it may be that what the theorist is faced with are equally acceptable content assignments. And if we neglect this option then we are in danger of letting irresolvable disputes about content proliferate.

# Bibliography

- Bach, K. (2007). 'Searle Against the World: How Can Experiences Find Their Objects?' in S. L. Tsohatzidis (ed.), *John Searle's Philosophy of Language: Force, Meaning and Mind* (pp. 64–78), Cambridge: Cambridge University Press.
- Bach, K. (2010). 'Getting a Thing into a Thought,' in R. Jeshion (ed.), *New Essays on Singular Thought* (pp. 39–64), Oxford: Oxford University Press.
- Baillargeon, R. (1987). 'Object Permanence in 3.5 and 4/5-Month-Old Infants,' *Developmental Psychology* 23: 655–64.
- Bakin, J.S., Nakayama, K., & Gilbert, C.D. (2000). 'Visual Responses in Monkey Areas V1 and V2 to Three-Dimensional Surface Configurations,' *Journal of Neuroscience* 20: 8188–98.
- Bayne, T. (2009). 'Perception and the Reach of Phenomenal Content,' *Philosophical Quarterly* 59(236): 385–404.
- Bermudez, J. (2000). 'Naturalized Sense Data,' *Philosophy and Phenomenological Research* 61(2): 353–74.
- Biederman, I. (1987). 'Recognition by Components: A Theory of Human Image Understanding,' *Psychological Review* 94: 115–47.
- Block, N. (2013). 'The Grain of Vision and the Grain of Attention,' *Thought* 1: 170–84.
- Block, N. Forthcoming. 'Seeing-As in the Light of Vision Science,' *Philosophy and Phenomenological Research*.
- Brewer, B. (2007). 'Perception and Its Objects,' *Philosophical Studies* 132: 87–97.
- Broad, C. D. (1925). *The Mind and Its Place in Nature*, London: Routledge and Kegan Paul.
- Brogaard, B. (2010). 'Strong Representationalism and Centered Content,' *Philosophical Studies* 151: 373–92.

- Brogaard, B. (2011). 'Centered Worlds and the Content of Perception,' in S. Hales (ed.), *Blackwell Companion to Relativism* (pp. 137–58), Oxford: Blackwell.
- Buckingham, G., Cant, J., & Goodale, M. (2009). 'Living in a Material World: How Visual Cues to Material Properties Affect the Way That We Lift Objects and Perceive Their Weight,' *Journal of Neurophysiology* 102: 3111–18.
- Burge, T. (2009). 'Five Theses on *De Re* States and Attitudes,' in J. Almog & P. Leonardi (eds.), *The Philosophy of David Kaplan* (pp. 246–316), Oxford: Oxford University Press.
- Burge, T. (2010). *Origins of Objectivity*, Oxford: Oxford University Press.
- Bushnell, B., Harding, P., Kosai, Y., & Pasupathy, A. (2011). 'Partial Occlusion Modulates Contour-Based Shape Encoding in Primate Area V4,' *Journal of Neuroscience* 31: 4012–24.
- Byrne, A. (2001). 'Intentionalism Defended,' *The Philosophical Review* 110: 199–240.
- Campbell, J. (2002). *Reference and Consciousness*, Oxford: Clarendon Press.
- Cappelen, H. and Lepore, E. (2005). *Insensitive Semantics*, Oxford: Blackwell.
- Carey, S. (2009). *The Origin of Concepts*, Oxford: Oxford University Press.
- Carey, D. P., Harvey, M., & Milner, A. D. (1996). 'Visuomotor Sensitivity for Shape and Orientation in a Patient with Visual Form Agnosia,' *Neuropsychologia* 34: 329–37.
- Chalmers, D. (2006). *Perception and the Fall from Eden*, in T. Gendler & J. Hawthorne (eds.), *Perceptual Experience* (pp. 49–125), Oxford: Clarendon Press.
- Clark, A. (2006). 'Attention & Inscrutability: A Commentary on John Campbell, *Reference and Consciousness*, for the Pacific APA Meeting, Pasadena, California, 2004,' *Philosophical Studies* 127: 167–93.
- Clarke, T. (1965). 'Seeing Surfaces and Physical Objects,' in M. Black (ed.), *Philosophy in America* (pp. 98–114), George Allen & Unwin Ltd.



- Creem, S. H. & Proffitt, D. R. (2001). 'Grasping Objects by Their Handles: A Necessary Interaction Between Cognition and Action,' *Journal of Experimental Psychology: Human Perception and Performance* 27(1): 218–28.
- Davies, M. (1992). 'Perceptual Content and Local Supervenience,' *Proceedings of the Aristotelian Society* 91: 21–45.
- DeRose, K. (2009). *The Case for Contextualism: Knowledge, Skepticism, and Context, Vol. 1*, Oxford: Oxford University Press.
- Desimone, R. & Duncan, J. (1995). 'Neural Mechanisms of Selective Visual Attention,' *Annual Review of Neuroscience* 18: 193–222.
- Devitt, M. (2008). 'Resurrecting Biological Essentialism,' *Philosophy of Science* 75: 344–82.
- Devitt, M. (2009). 'Biological Realisms,' in H. Dyke (ed.), *From Truth to Reality: New Essays in Logic and Metaphysics*, London: Routledge and Kegan Paul.
- Devitt, M. (2011). 'Natural Kinds and Biological Realisms,' in M. O'Rourke, J. Keim Campbell & M. H. Slater (eds.), *Carving Nature at its Joints*, Cambridge, MA.: MIT Press.
- Dickie, I. (2010). 'We Are Acquainted With Ordinary Things,' in R. Jeshion (ed.), *New Essays on Singular Thought* (pp. 213–45), Oxford: Oxford University Press.
- Dickie, I. (2011). 'Visual Attention Fixes Demonstrative Reference By Eliminating Referential Luck,' in C. Mole, D. Smithies, & W. Wu (eds.), *Attention: Philosophical and Psychological Essays* (pp. 292–322), Oxford: Oxford University Press.
- Dretske, F. (1969). *Seeing and Knowing*, Chicago: Chicago University Press.
- Dretske, F. (1986). 'Misrepresentation,' in Radu Bogdan (ed.), *Belief: Form, content and function* (pp. 17–36), New York: Oxford.
- Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA.: MIT Press.

- Dretske, F. (1990). 'Seeing, Believing, and Knowing,' in D. Osherson, S. Kosslyn, and J. Hollerbach (eds.), *Visual Cognition and Action* vol. II (pp. 129–48), Cambridge, MA.: MIT Press.
- Driver, J., Davis, G., Russell, C., Turatto, M., & Freeman, E. (2001). 'Segmentation, Attention and Phenomenal Visual Objects,' *Cognition* 80: 61–95.
- Dummett, M. (1973). *Frege: Philosophy of Language*, London: Duckworth.
- Dummett, M. (1981). *The Interpretation of Frege's Philosophy*, London Duckworth.
- Egly, R., Driver, J., & Rafal, R. D. (1994). 'Shifting Visual Attention Between Objects and Locations: Evidence From Normal and Parietal Lesion Subjects,' *Journal of Experimental Psychology: General* 123: 161–77.
- Evans, G. (1975). 'Identity and Predication,' in *Collected Papers*, Oxford: Clarendon Press.
- Field, H. (1973). 'Theory Change and the Indeterminacy of Reference,' *Journal of Philosophy* 70(14): 462–81.
- Fodor, J. (1994). *The Elm and the Expert*. Cambridge, MA.: MIT Press.
- Fodor, J. (1990). 'A Theory of Content,' in *A Theory of Content and Other Essays*, Cambridge, MA: MIT/Bradford Press.
- Fodor, J. (2007). 'Revenge of the Given,' in B. McLaughlin & J. Cohen (eds.), *Contemporary Debates in Philosophy of Mind* (pp. 105–16), Oxford: Blackwell.
- Frisby, J., & Stone, J. (2010). *Seeing: The Computational Approach to Biological Vision (Second Edition)*, Cambridge, MA.: MIT Press.
- Gates, G. (1996). 'The Price of Information,' *Synthese* 107(3): 325–47.
- Godfrey-Smith, P. (2005). 'Folk-Psychology as a Model,' *Philosopher's Imprint* Vol. 5, No. 6: 1–16.
- Godfrey-Smith, P. (2006). 'The Strategy of Model-Based Science,' *Biology and Philosophy* 21: 725–40.

- Grice, H. P. (1961). 'The Causal Theory of Perception,' in *Proceedings of the Aristotelian Society* 35: 121–68.
- Grice, H. P. (1975). 'Logic and Conversation,' in P. Cole & J. Morgan (eds.), *Syntax and Semantics, 3: Speech Acts* (pp. 41–58), New York: Academic Press. Reprinted in H. P. Grice (ed.), *Studies in the Way of Words* (pp. 22–40), Cambridge, MA.: Harvard University Press (1989).
- Harman, G. (1990). 'The Intrinsic Quality of Experience,' *Philosophical Perspectives* 4, ed. J. Tomberlin (pp. 31–52), Northridge: Ridgeview Publishing Company.
- Hawthorne, J. (2004). *Knowledge and Lotteries*, Oxford: Oxford University Press.
- Healey, C. & Enns, J. (2012). 'Attention and Visual Memory in Visualization and Computer Graphics,' *IEE Transactions on Visualization and Computer Graphics* Vol. 18, No. 7: 1170–1188.
- Heck, R. G. (2000). 'Nonconceptual Content and the Space of Reasons,' *Philosophical Review* 109: 483–523.
- Heck, R. G. (2007). 'Are There Different Kinds of Content?' in J. Cohen and B. McLaughlin (eds.), *Contemporary Debates in the Philosophy of Mind* (pp. 117–38), Oxford: Blackwell.
- Hill, C. (2009). *Consciousness*, Cambridge: Cambridge University Press.
- Howe, P. D., Cohen, M. A., Pinto, Y. & Horowitz, T. H. (2010). 'Distinguishing Between Parallel and Serial Accounts of Multiple Object Tracking,' *Journal of Vision* 10(8): 11, 1–13.
- Huemer, M. (2008). 'When is Parsimony a Virtue?' *The Philosophical Quarterly* 59(235): 216–36.
- Huntley-Fenner, G., Carey, S., & Salimando, A. (2002). 'Objects are Individuals but Stuff Doesn't Count: Perceived Rigidity and Cohesiveness Influence in Infants' Representation of Small Numbers of Discrete Entities,' *Cognition* 85: 203–21.

- Jackson, F. (1977). *Perception: A Representative Theory*, New York: Cambridge University Press.
- Jackson, F. (2012a). 'Perception, Representation, Language,' *The Amherst Lecture in Philosophy* 7 (pp. 1–17): <<http://www.amherstlecture.org/jackson2012/>>
- Jackson, F. (2012b). 'Michael Tye on Perceptual Content,' *Philosophy and Phenomenological Research* 84(1): 199–205.
- James, T. W., Culham, J., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2003). 'Ventral Occipital Lesions Impair Object Recognition but Not Object-Directed Grasping: An fMRI Study,' *Brain* 126: 2463–75.
- Johnston, M. (2004). 'The Obscure Object of Hallucination,' *Philosophical Studies* 120: 113–83.
- Johnston, M. (2006). 'Better Than Mere Knowledge? The Function of Sensory Awareness,' in T. S. Gendler and J. Hawthorne (eds.), *Perceptual Experience* (pp. 260–90), Oxford: Oxford University Press.
- Kastner, S. & Ungerleider, G. (2000). 'Mechanisms of Visual Attention in the Human Cortex,' *Annual Review of Neuroscience* 23: 315–41.
- Kellman, P. & Spelke, E. S. (1983). 'Perception of Partly Occluded Objects in Infancy,' *Cognitive Psychology* 15: 483–524.
- Kennedy, C. (1999). *Projecting the Adjective: The Syntax and Semantics of Gradability and Comparison*, New York: Routledge.
- Kriegel, U. (2009). *Subjective Consciousness: A Self-Representational Theory*, New York: Oxford University Press.
- Kripke, S. (1980). *Naming and Necessity*, Oxford: Basil Blackwell.
- Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2014). 'A Conceptual Framework of Computations in Mid-Level Vision,' *Frontiers in Computational Neuroscience* 8: 158 doi: 10.3389/fncom.2014.00158

- Lewis, D. (1980). 'Veridical Hallucination and Prosthetic Vision,' *Australasian Journal of Philosophy* 58(3): 239–49.
- Marr, D. & Nishihara, H. (1978). 'Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes,' *Proceedings of the Royal Society of London* 200: 269–94.
- Martin, M. G. F. (2002). 'Particular Thoughts and Singular Thoughts,' in A. O'Hear (ed.), *Logic, Thought and Language* (pp. 173–214), Cambridge: Cambridge University Press.
- McDowell, J. (1982). 'Criteria, Defeasibility, and Knowledge,' *Proceedings of the British Academy*, 68: 455–79.
- McGinn, C. (1982). *The Character of Mind*, Oxford: Oxford University Press.
- McLaughlin, B. (1984). 'Perception, Causation, and Supervenience,' *Midwest Studies in Philosophy* IX: 569–91.
- Mehta, N. (2014). 'The Limited Role of Particulars in Phenomenal Experience,' *The Journal of Philosophy* 111(6): 311–31.
- Millar, A. (1991). *Reasons and Experience*, Oxford: Clarendon Press.
- Millikan, R. (1984). *Language, Thought and Other Biological Categories*, Cambridge, MA.: MIT Press.
- Millikan, R. (2004). *Varieties of Meaning*, Cambridge, MA.: MIT Press.
- Milner, A. D. & Goodale, M. A. (1995). *The Visual Brain in Action*, Oxford: Oxford University Press.
- Milner, A. D. & Goodale, M. A. (2006). *The Visual Brain in Action* (2<sup>nd</sup> edition), Oxford: Oxford University Press.
- Milner, A. D. & Goodale, M. A. (2008). 'Two Visual Systems Re-Viewed,' *Neuropsychologia* 46: 774–85.

- Moore, G. E. (1953). *Some Main Problems of Philosophy*, London: George, Allen and Unwin.
- Nakayama, K., He, Z. J., & Shimojo, S. (1995). 'Visual Surface Representation: A Critical Link between Lower-Level and Higher-Level Vision,' in S. M. Kosslyn & D.N. Osherson (eds.), *An Invitation to Cognitive Science* (2<sup>nd</sup> edition) (pp. 1–70), Cambridge, MA.: MIT Press.
- Nanay, B. (2010a). 'Perception and Imagination: Amodal Perception as Mental Imagery,' *Philosophical Studies* 150: 239–54.
- Nanay, B. (2010b). 'A Modal Theory of Function,' *Journal of Philosophy* 107: 412–31
- Nanay, B. (2011). 'Function, Modality and Mental Content,' *Journal of Mind and Behavior*, 32: 84–7.
- Nanay, B. (2012). 'Function Attribution Depends on the Explanatory Context,' *Journal of Philosophy* 109: 623–7.
- Nanay, B. (2014). 'Empirical Problems with Anti-Representationalism,' in B. Brogaard (ed.), *Does Perception Have Content?* (pp. 39–50), Oxford: Oxford University Press.
- Nanay, B. (2015). 'The Representationalism versus Relationism Debate: Explanatory Contextualism about Perception,' *European Journal of Philosophy* 23(2): 321–36.
- Neta, R. (2007). 'Contextualism and a Puzzle About Seeing,' *Philosophical Studies* 134(1): 53–63.
- Pautz, A. (2009). 'What are the Contents of Experiences?' *Philosophical Quarterly* 59(236): 483–507.
- Phillips, B. (2012). 'Modified Occam's Razor,' *Australasian Journal of Philosophy* 90(2): 371–82.
- Phillips, B. (2014). 'Indirect Representation and the Self-Representational Theory of Consciousness,' *Philosophical Studies* 167: 273–90.
- Prinz, J. (2002). *Furnishing the Mind: Concepts and their Perceptual Basis*, Cambridge,

- MA.: MIT Press.
- Prinz, J. (2013). 'Siegel's Get Rich Quick Scheme,' *Philosophical Studies* 163(3): 827–35.
- Putnam, H. (1975). "The Meaning of 'Meaning,'" *Minnesota Studies in the Philosophy of Science* 7: 215–71.
- Pylyshyn, Z. W. & Annan, V. (2006). 'Dynamics of Target Selection in Multiple Object Tracking (MOT),' *Spatial Vision* 19: 485–504.
- Pylyshyn, Z. W. (2001). 'Visual Indexes, Preconceptual Objects, and Situated Vision,' *Cognition*, 80 (1/2): 127–58.
- Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*, Cambridge, MA.: MIT Press.
- Pylyshyn, Z. W. (2006). 'Some Puzzling Findings in Multiple Object Tracking (MOT): II. Inhibition of Moving Nontargets,' *Visual Cognition* 14(2): 175–198
- Pylyshyn, Z. W. (2007). *Things and Places: How the Mind Connects with the World*, Cambridge, MA.: MIT Press.
- Quine, W. V. O. (1960). *Word and Object*, Cambridge, MA.: MIT Press.
- Quine, W. V. O. (1969). *From a Logical Point of View*, Cambridge, MA.: Harvard University Press.
- Raftopoulos, A. (2009). *Cognition and Perception: How do Psychology and Neural Science Inform Philosophy?* Cambridge, MA.: MIT Press.
- Rensink, R. A. & Enns, J. T. (1998). 'Early Completion of Occluded Objects,' *Vision Research* 38: 2489–2505.
- Rosenthal, D. M. (1991). 'The Independence of Consciousness and Sensory Quality,' in E. Villanueva (ed.), *Consciousness: Philosophical Issues* (pp. 15–36), Vol. 1, Atascadero, CA: Ridgeview Publishing Company.

- Rosenthal, D. M. (1999). 'The Colors and Shapes of Visual Experiences,' in D. Fisette (ed.), *Consciousness and Intentionality: Models and Modalities of Attribution* (pp. 95–118), Dordrecht: Kluwer Academic Publishers.
- Rosenthal, D. M. (2005). *Consciousness and Mind*, Oxford: Clarendon Press.
- Rosenthal, D. M. (2010). 'How to Think About Mental Qualities,' *Phil. Issues* 20: 368–93. doi: 10.1111/j.1533-6077.2010.00190.x
- Rosenthal, D. M. (in press). 'Quality Spaces, Relocation, and Grain,' in J. R. O'Shea (ed.), *Sellars and His Legacy*, Oxford: Oxford University Press.
- Sato, T., Uchida, G., Lescroart, M., Kitazono, J., Okada, M., & Tanifuji, M. (2013). 'Object Representation in Inferior Temporal Cortex Is Organized Hierarchically in a Mosaic-Like Structure,' *The Journal of Neuroscience* 33(42): 16642–16656.
- Schaffer, J. (2006). 'The Irrelevance of the Subject,' *Philosophical Studies* 127: 87–107.
- Schellenberg, S. (2010). 'The Particularity and Phenomenology of Perceptual Experience,' *Philosophical Studies* 149: 19–48.
- Schellenberg, S. (2011). 'Perceptual Content Defended,' *Noûs* 45(4): 714–50.
- Schenk, T. & McIntosh, R. D. (2010). 'Do We Have Independent Visual Streams for Vision and Action?' *Cognitive Neuroscience* 1: 52–62.
- Scholl, B. (2001). *Objects and Attention*, Cambridge MA.: MIT Press.
- Scholl, B. J. & Pylyshyn, Z. W. (1999). 'Tracking Multiple Items through Occlusion: Clues to Visual Objecthood,' *Cognitive Psychology* 38: 259–90.
- Scholl, B., Pylyshyn, Z. W., & Feldman, J. (2001). 'What is a Visual Object? Evidence From Target Merging in Multiple Object Tracking,' *Cognition* 80: 159–77.
- Searle, J. (1983). *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press.
- Searle, J. (1987). 'Indeterminacy, Empiricism, and the First Person,' *Journal of Philosophy* 84(3): 123–46.



- Siegel, S. (2006). 'How Does Phenomenology Constrain Object-Seeing?' *Australasian Journal of Philosophy*, 84(3): 429–41.
- Siegel, S. (2011). *The Contents of Visual Experience*, Oxford: Oxford University Press.
- Siewert, C. (1998). *The Significance of Consciousness*, Princeton: Princeton University Press.
- Soteriou, M. (2000). 'The Particularity of Visual Perception,' *European Journal of Philosophy* 8(2): 173–89.
- Speaks, J. (2005). 'Is There a Problem about Nonconceptual Content?' *Philosophical Review* 114: 359–98.
- Spelke, E. (1990). 'Principles of Object Perception,' *Cognitive Science* 14: 29–56.
- Spelke, E., Kestenbaum, R., Simons, D., & Wein, D. (1995). 'Spatio-Temporal Continuity, Smoothness of Motion and Object Identity in Infancy,' *British Journal of Developmental Psychology* 13: 113–42.
- Stanford, P. K. (1995). 'For Pluralism and Against Realism about Species,' *Philosophy of Science* 62: 70–91.
- Stanley, J. (2004). 'On the Linguistic Basis for Contextualism,' *Philosophical Studies* 119: 119–146.
- Stanley, J. (2005). *Knowledge and Practical Interests*, Oxford: Oxford University Press.
- Strawson, P. F. (1959). *Individuals*, London: Methuen.
- Sugita, Y. (1999). 'Grouping of Image Fragments in Primary Visual Cortex,' *Nature* 401: 269–72.
- Travis, C. (2004). 'The Silence of the Senses,' *Mind* 113: 57–94.
- Treisman, A. (1985). 'Preattentive Processing in Vision,' *Computer Vision, Graphics and Image Processing* 31: 156–77.
- Tse, P. (1999). 'Volume Completion,' *Cognitive Psychology* 39: 37–68.
- Tye, M. (2009). 'The Admissible Contents of Visual Experience,' *The Philosophical*

*Quarterly* 59(236): 541–62.

van Lier, R. & Gerbino, W. (2015). ‘Perceptual Completions,’ in Wagemans, J. (ed.), *The Oxford Handbook of Perceptual Organization*, New York: Oxford University Press.

van Lier, R. & Wagemans, J. (1999). ‘From Images to Objects: Global and Local Completions of Self-Occluded Parts,’ *Journal of Experimental Psychology: Human Perception and Performance* 25(6): 1721–41.

vanMarle, K. & Scholl, B. (2003). ‘Attentive Tracking of Objects and Substances,’ *Psychological Science* 14(5): 498–504.

Wilcock, J. (ms). ‘Two Failures to Scrutinize Reference.’

Wu, W. (2008). ‘Visual Attention, Conceptual Content, and Doing it Right,’ *Mind* 117 (468): 1003–33.

Yantis, S. (1998). ‘Objects, Attention, and Perceptual Experience,’ in R. Wright (ed.), *Visual Attention* (pp. 187–214), Oxford: Oxford University Press.