Synthese (2017) 194:2949–2971
DOI 10.1007/s11229-016-1083-3

Inscrutability and visual objects

Ben Phillips[1]

Received: 30 July 2015 / Accepted: 26 March 2016 / Published online: 11 April 2016

Abstract The thesis that the visual system represents objects has garnered empirical support from a variety of sources in recent decades. But what kinds of things qualify as "objects" in the relevant sense? Are they ordinary three-dimensional bodies? Are they the facing surfaces of three-dimensional bodies? I argue that there is no fact of the matter: what we have are equally acceptable ways of assigning extensions to the relevant visual states. The view I defend bears obvious similarities to Quine's thesis that linguistic reference is inscrutable. Importantly, though, I argue that even if Quine was wrong about inscrutability as a thesis about language and thought, the case for the inscrutability of visual reference remains strong.

Keywords Perception · Inscrutability · Visual objects · Indeterminacy · Visual reference · Multiple-object tracking

1 Introduction

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 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 perceive objects as having occluded parts—seems to occur in early and mid-level vision, before object recognition takes place (Rensink and Enns

✉ Ben Phillips
ben.s.phillips@gmail.com

[1] The Graduate Center, City University of New York, 365 Fifth Ave, New York, NY 10016, USA

1998; Sugita 1999). The environmental particulars that are represented by the states that participate in these visual processes are often referred to as "visual objects."

In this paper, I address the following questions: What exactly are visual objects? Are they ordinary 3D objects? Are they the facing surfaces of ordinary 3D objects? I will argue that there is just no fact of the matter as to whether the states in question represent 3D objects or just their facing surfaces. I will do 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, he argues that the term 'rabbit' is referentially indeterminate between the set of rabbits and the set of undetached-rabbit-parts. However, similar considerations can be used to argue that there is no fact of the matter as to whether the visual states that mediate object-tracking and amodal completion pick out ordinary objects or just their facing surfaces: after all, the facing surfaces of objects are undetached parts of them.[1] Other than answering an important and traditional question about the contents of visual perception, the indeterminacy thesis I seek to defend is significant because it undermines the seductive view that perceptual content is the root of all content determinacy.[2]

The structure of the paper is as follows. In Sect. 2, I review empirical evidence for the claim that we visually represent objects. In Sect. 3, I argue that the data do not adjudicate between the view that these objects are ordinary 3D bodies, and the alternative according to which they are the facing surfaces of ordinary 3D bodies. In Sect. 4, I address different versions of 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 consequences for the nature of object-seeing.

2 Evidence for object-representations in vision

In this section, I review what I take to be the most compelling pieces of evidence for the view that vision houses object-representations. In subsequent sections, I take up the issue of just what counts as an "object" in the relevant sense.

[1] I will use the term 'facing surface' to refer to the part of an object from which light is reflected onto the subject's retinas. According to this usage, there is no such thing as a partially occluded facing surface. I will use the term 'front surface' to refer to the part of an object that would reflect light onto the subject's retinas were there no occluding objects. Finally, I will use the term 'facing portion' more broadly to include both the facing surfaces of objects and the 3D object-parts of which these facing surfaces are parts. For instance, if my hand is protruding from behind the curtain then it counts as my facing portion, as does its facing surface (e.g. my palm).

[2] For examples of the view that perceptual content is the root of all content determinacy, see Burge (2010), Pautz (2013), and Peacocke (2014).

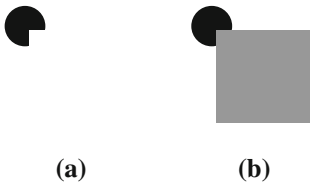


Fig. 1 Notched versus partially occluded circle

Fig. 2 Visual search task: version 1

Fig. 3 Visual search task: version 2

2.1 Amodal completion and visual search tasks

Consider Fig. 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*.

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 Figs. 2 and 3.

What Rensink and Enns found is that it is much harder to locate the notched circle in 2 than it is in 3. They found that in 3, the task was 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 yielded 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.[3] When you examine Fig. 3, early visual processes

[3] The term "pre-attentive" must be used with caution. Single cell recordings in animals (Desimone and Duncan 1995) and functional imaging in humans (Kastner and Ungerleider 2000) have shown that many

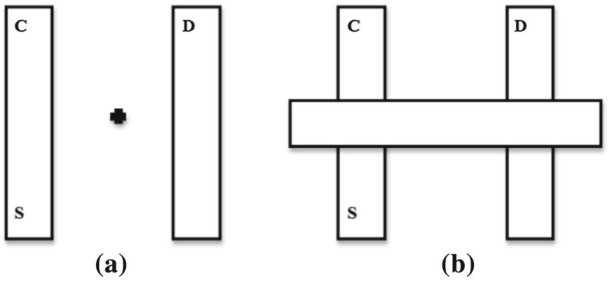


Fig. 4 The automatic spread of attention

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 Fig. 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 Fig. 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. 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.

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 Fig. 4.

In a series of trials, Egly et al. (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 Fig. 4b). Again, the most straightforward explanation is that objects are represented by the visual system pre-attentively—or at least, relatively early on in processing—before any of them are selectively attended to for further processing.

Footnote 3 continued
levels of processing (including processing within primary visual cortex) can be modulated by top-down selective attention.

Fig. 5 Partially occluded bar

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 macaques are sensitive to facts concerning occlusion. Subjects were presented with two vertical bars, separated by a grey patch (as in Fig. 5).

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, p. 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.[4]

2.4 Volume completion

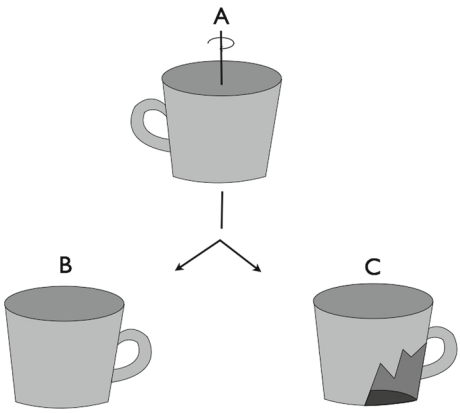
The studies I have described so far center mainly on surface completion, but there are compelling reasons for thinking that volume completion is a visual process as well. For instance, when you look at a bowling ball, its front surface occludes its rear surface. Nonetheless, from a phenomenological point of view, it is plausible that you visually experience the ball's sphericality—a volumetric property that requires a form of amodal completion over and above 2D surface completion.[5]

Another reason for construing volume completion as a decidedly perceptual phenomenon concerns the fact that we complete the rear surfaces of self-occluding objects in a way that is impervious to our beliefs about them. For instance, consider Fig. 6.

[4] For a more recent study utilizing single-cell recordings, see Bushnell et al. (2011). For a recent brain imaging study, see van Lier and Gerbino (2014).

[5] See Nakayama et al. (1995) for an influential treatment of surface completion. For some influential treatments of volume completion, see Tse (1999) and van Lier and Wagemans (1999).

Fig. 6 Volume completion



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.[6]

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. 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.[7] But successfully picking up ordinary objects requires representations of their volumetric properties (see Sect. 5): 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.[8]

2.5 Multiple-object tracking

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 asked to track these targets as all 8 figures move along independent trajectories. At some point (5–10 s later) the motion stops and the subject is asked to indicate which figures were targets. Pylyshyn (2003, pp. 223–232) has found that subjects are able to track 4–5 objects with an efficiency characteristic of parallel processing.

[6] See Nanay (2010) for an extended version of this argument.

[7] See James et al. (2003).

[8] See Tse (1999, p. 65) for a similar argument.

However, beyond 5, the capacity falls off sharply, with a profile characteristic of serial processing.[9]

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, before object recognition occurs.

Equally important as the fact that MOT appears to be a visual process is the fact that as long as 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, and shape, 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.[10] Tracking also breaks down if the target fails to move along a spatiotemporally continuous trajectory. Finally, experiments have shown that subjects cannot track easily defined *parts* of objects. Scholl et al. (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.

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 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?

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 one 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 the relevant data. Consequently, there is just no fact of the matter as to which assignment is correct: reference is inscrutable. For

[9] See Howe et al. (2010) for a study that directly tests, and supports, the claim that MOT occurs via parallel processing.

[10] See Scholl and Pylyshyn (1999), as well as van Marle and Scholl (2003).

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.

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 unacceptable 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, p. 126).

Now, my concern 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. Rather, I want to go about arguing that, as far as the relevant data are concerned, *visual* reference is indeterminate 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.

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 Fig. 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* (where this facing portion has the same shape as a notched circle). In explaining why the subject finds it harder to locate the target in Fig. 2, the key point is that her visual system does not represent it *as a notched circle*. 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 the Egly et al. (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 portions of objects: that is to say, we can construe the subject's attention as spreading out over the facing surfaces of objects, such as the bars in Fig. 4.[11]

What goes for surface completion goes for volume completion too. For instance, consider those versions of Rensink and Enns' visual search task in which targets and distractors are 3D bodies (not 2D squares and circles on a computer screen). The

[11] It is worth emphasizing that both 3D bodies and their front surfaces have facing portions in their own right. Thus, even if surface completion occurs prior to volume completion in order of visual processing, representations at both levels of processing exhibit the kind of indeterminacy that I'm positing.

analogue of the notched circle is a notched sphere with a 3D chunk taken out of it, and the competing reference-schemes that each allow us to explain the subject's performance on the relevant search task will be along the following lines (with obvious simplifications): *partially occluded sphere (of such-and-such dimensions)* and *facing portion (of such-and-such dimensions) of partially occluded sphere (of such-and-such dimensions)*.[12]

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.

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 time, but not in virtue of encoding their properties. Pylyshyn's visual indexes are thus akin to pure demonstratives (e.g. "that"). 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, p. 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

[12] In construing volume completion as a visual phenomenon, I'm not conceding that the states involved determinately refer to *both* 3D objects and their facing surfaces. For instance, consider a case of spherical volume completion. According to my view, the extension of the resulting state is indeterminate between *being a sphere* and *being the facing surface of a sphere*: two properties that are clearly not co-extensive. In the same way, claiming that one cannot refer to an undetached-rabbit-part without also determinately referring to a rabbit would not go any way towards undermining Quine's inscrutability thesis: it would simply beg the question against Quine. See Gates (1996, p. 331) for further discussion of this issue.

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—picture his finger running over the surface of the ball in the same way that a potter's fingers run over the surface of a clay pot as it spins on the wheel.

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.

One might worry that the account of object-tracking I have just provided will only work if we construe the visual index in question as picking out successive facing portions of the same object, as well as representing them *as portions of the same persisting object*. After all, what if the currently indexed portion were to break apart from the object being tracked? Studies have shown that in this scenario, either the subject will continue to track the object itself (not the portion that broke off), or, tracking will break down altogether. Doesn't this show that the reference-scheme offered above will only allow us to explain object-tracking if the facing portions that are successively being picked out are being represented *as portions of the same object*?[13]

It would be problematic for the indeterminacy view I'm defending if it forced us to answer this question in the affirmative, for on the standard view—according to which visual indexes determinately refer to tracked objects, not their facing portions—there does not appear to be any need to posit explicit representations of numerical identity. To be sure, subjects can track objects that disappear and then reappear from behind occluders.[14] However, to claim that this requires the subject to explicitly represent the object that reappears from behind the occluder *as numerically identical to the object that disappeared behind it* would be to over-intellectualize object-tracking. More specifically, to do so would be to conflate the principles according to which the object-tracking system assigns and maintains visual indexes, with the contents of the representations involved. The object-tracking system is designed to track spatiotemporally continuous bodies, even when those bodies are temporarily occluded: to posit explicit representations of numerical identity in addition to these operating principles would therefore be to posit contents that are doing no explanatory work.[15]

What this means is that if the worry canvassed above is a good one—that is to say, tracking an object by indexing its successive facing portions would require explicitly

[13] Thank you to an anonymous referee for raising this concern.

[14] See Kellman and Spelke (1983); and Spelke (1990).

[15] See Burge (2011, p. 125) for a similar point.

representing those portions *as portions of the same object*—we would have a reason to favor the standard view on grounds of parsimony. Needless to say, it is clear that the indeterminacy-theorist can defuse this worry by simply invoking the distinction—invoked above—between principles of operation and the explicit representations that are tokened as a result. By definition, facing portions go wherever the objects of which they are portions go. Thus, as long as the object-tracking system operates in such a fashion that a visual index is only assigned to the successive facing portions of a single *persisting* object, we have all that we need to explain object-tracking behavior.[16]

3.3.1 Visual indexes with attributive elements

I have just argued that on the reference-scheme according to which a visual index refers to successive facing portions of a tracked object, these facing portions needn't be represented *as portions of the same object*. But do *any* object-features need to be encoded for visual indexes to do their work, and if so, are these feature-representations compatible with the indeterminacy thesis on offer?

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

Elsewhere, he elaborates as follows:

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

For instance, simplifying somewhat, 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 between 3D bodies and their facing surfaces. For instance, according to the latter reference scheme, if I'm tracking a rotating body then V will be along the lines of *THAT FACING-PORTION-OF-COHESIVE-BOUNDED-BODY*—a complex demonstrative that picks out different facing portions of *the very same body* as it rotates.

[16] In any case, even if one *were* convinced that tracking through occlusion requires explicit representations of facts concerning numerical identity, the inscrutability thesis I'm defending could accommodate this. For instead of construing the subject's visual state as representing the object in question as being numerically identical to a previously indexed one, we could construe her as representing its current facing portion *as belonging* to the very same object whose facing portion was visible a moment ago. Quine (1969, p. 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?"

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 is a thin one. All that is required for a particular to count as an object—in this *thin* sense of the term—is for it to be a cohesive and bounded 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 so there is nothing to choose between them as far as fixing visual reference goes.[17,18]

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 as well?

As I see it, there are two main versions of this worry. According to the first, there are arguments for the view that linguistic expressions and thoughts have determinate extensions: arguments that apply with equal force to the perceptual case. According to the second, 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.

4.1 Fodor's argument for the determinacy of linguistic reference

In an influential reply to Quine's inscrutability challenge, Fodor (1994, p. 65) asks us to consider a figure of the form shown above (Fig. 7).

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

[17] Spelke (1990) develops a conception of objecthood that fits with what I'm calling the "thin" conception. One might think that the visual system also represents objects as *solid*—see Burge (2010, pp. 465–471) for a discussion of this question—but this is entirely compatible with the inscrutability of visual reference: on the deviant interpretation, the visual system picks out the facing surfaces of *solid*, cohesive, bounded, and spatiotemporally continuous bodies and represents them as such.

[18] As an anonymous reviewer has pointed out, we often perceive particulars that don't have facing *surfaces*: e.g. holograms, shadows, rainbows, and rays of light. Nonetheless, these entities do have facing *portions*. For instance, an appropriately positioned tree can occlude the bottom half of your shadow; a mountain can occlude a portion of a rainbow; and so on. The parts that are visible in these cases qualify as facing portions, even if they don't qualify as facing *surfaces*. Perhaps it is possible to perceive an object with no occluded parts—the object and its facing portion are thus one and the same. However, the existence of degenerate cases like this would do nothing to undermine my thesis that in ordinary cases, the state in question is indeterminate between a three-dimensional body and its facing surface. In order to bring this point out, notice that Quine's thesis that 'rabbit' is indeterminate between rabbits and undetached-rabbit-parts would not be undermined, in the slightest, if it turned out that there are simples with no undetached parts (e.g. fundamental entities, such as strings).

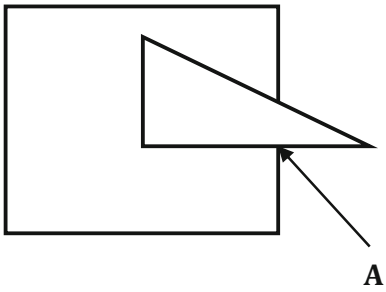


Fig. 7 Intersection of a triangle and a square

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 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, pp. 74–77), his solution to Quine's challenge involves a rejection of purely informational theories of content, for it requires us to construe the inferential roles of the conjunction symbol as (partly) fixing the extensions of our predicates. For instance, the extension of 'is a square' is fixed, in part, by the fact that ordinary speakers are disposed to accept the inference from "*A* is a square" and "*A* is an *F*" to "*A* is a square and an *F*," for arbitrary predicate 'is an *F*.'

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 concept *CUBE*. According to the standard interpretation, this concept picks out the set of cubes; whereas, on the deviant interpretation, it picks out the set of facing-surfaces-of-cubes. Now, the concept *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 is orientated in the right way, the thought *THAT IS BOTH A CUBE AND A SQUARE* comes out true on the interpretation according to which *CUBE* has the set of facing-surfaces-of-cubes as its extension. We should therefore draw the conclusion that the concept *CUBE* does not have the set of facing-surfaces-of-cubes as its extension—or so the argument goes.

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 inscrutable, I will not take a stance on whether it works

in the case of language and thought.[19] Rather, I will simply assume—for the sake of argument—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 that the attributive elements tokened in vision are the same ones that figure downstream in thought. Unless this has been established, we do not have the leverage we need to infer from the fact that the subject dissents (in thought) from *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 between objects and their facing surfaces; whereas, those states tokened downstream in cognition have determinate extensions. And the view that indeterminate representations in vision give rise to determinate ones 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 serious work to argue against.

4.3.1 Evans' strategy and visual reference

Analogous remarks apply to Evans' (1975) argument against Quine's inscrutability thesis.[20] The argument involves compound predicates of the form '*FG*,' the application of which ordinary speakers will dissent from, despite the fact that they assent to the application of '*F*' and '*G*' separately. For instance, Evans provides the example of the compound 'white rabbit.' If one is in the presence of a brown rabbit with a white foot, one will dissent from "There is a white rabbit." On the other hand, one will assent to both "There is a rabbit" and "There is whiteness." In order for "There is a white rabbit" to be assented to the whiteness must be roughly distributed over the entire surface of the rabbit.

According to Evans, these considerations rule out Quine's deviant reference-schemes because they are not sensitive to the boundaries of ordinary objects, such as rabbits. For instance, 'rabbit' cannot refer to the set of undetached-rabbit-parts because this would get the truth-conditions of "There is a white rabbit" wrong: one would clearly dissent from this sentence in the presence of a brown rabbit with a white foot, despite the fact that there is a white undetached-rabbit-part present. Evans' diagnosis is that the *F*-feature in question must "be distributed in a characteristic way in relation to the boundaries of a *single* object whose presence prompts assent or dissent to relevant queried *G*-terms." (1975, p. 351)

[19] See Gates (1996) for a persuasive argument that Fodor's argument is unsound.

[20] Fodor (1994, p. 123) construes his own objection as a development of the insights found in Evans' article.

Burge (2010, pp. 216–234) provides persuasive reasons for rejecting Evans' argument on its own terms. At any rate, notice that what I said above about Fodor's strategy applies with equal force to Evans': that is to say, even if it works as an argument that reference is determinate at the level of language and thought, to draw like conclusions about the determinacy of *visual* reference would be to seriously overreach. In order to draw like conclusions about the determinacy of visual reference, Evans would have to show that the attributive elements tokened in vision are the same ones that are tokened downstream in thought, and needless to say, nothing he says goes any way towards establishing this.

Furthermore, notice that the various reference-schemes I have construed as equally legitimate when it comes to MOT, amodal completion, and so on, are all schemes that exhibit the requisite sensitivity to the boundaries of objects and the ways in which features are distributed over them. In order to illustrate this point, consider a perceptual analogue of Evans' 'white rabbit' example. Suppose you are faced with a brown sphere that has a white fleck on its facing surface. Let's stipulate that your visual state is accurate.

Now, if your visual state were, in fact, representing the presence of a white sphere, or, *the facing surface* of a white sphere then it would obviously count as inaccurate. But notice that the inaccuracy of your visual state in this case is something that could easily be accommodated on the alternative reference-schemes we have been considering. For instance, on the scheme according to which your visual state picks out the facing surface of the sphere and represents it *as the facing surface of a white sphere*, your state clearly comes out as inaccurate: the facing surface in front of you is part of a brown sphere, not a white one.

4.4 The argument from demonstrative thought

According to the second argument from cognitive consumption, perceptual demonstrative thoughts inherit their referents from the perceptual states that give rise to them.[21] 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 true that my view on the indeterminacy of visual reference is incompatible with the thesis that the referents of perceptual demonstrative thoughts are inherited *exclusively* from those perceptual states upstream that mediate amodal completion and object tracking. Be that as it may, it is far from clear why a visual state with indeterminate referents could not give rise to a demonstrative thought with determinate ones.

[21] For example, see Martin (2002).

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. In order to see why this is the case, recall that according to Fodor, no purely informational theory of content can secure referential determinacy: necessarily, terms 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 rich array of inferential transitions that they participate in downstream of MOT—namely, inferences involving predicate conjunction and other logico-syntactic apparatus.

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 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.

5.1 Object-manipulation and visual inscrutability

Let's work through a simple example: you are thirsty, and so you pick up the cup that is sitting on the table. In order to explain how you performed this action, must we construe the *visual* object-representations that mediated it as having determinate extensions? In order to answer this question, 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 attend to the cup, and recognize it as such.

The second step involves the formulation of an appropriate action-plan. Among other things, you must select which hand to use; an appropriate type of grip; and an appropriate type of posture. The selection of initial grip and lift forces are informed by past experiences with objects of the relevant visual appearance (Buckingham et al. 2009). For instance, your past experiences with cups, along with your recognition of the object *as a cup*, will aid you in selecting both an appropriate type of grip and an appropriate lift force (given your estimate of the cup's 3D shape and weight).

Already, at the planning stage, we have the makings of a challenge to the visual inscrutability 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 Sect. 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 et al. 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 the subject to form an action-plan that is sensitive to its 3D-shape. The key point here 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*.[22]

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 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; however, there is significant controversy regarding just how independent it is from object recognition and planning, which are ventral stream processes. According to Milner and Goodale's influential model (2006), 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 selec-

[22] Once again, it is important not to lose sight of the fact that *being a body of such-and-such a 3D-shape* and *being the facing surface of a body of such-and-such a 3D-shape* are not co-extensive properties.

tion 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.[23] For our purposes, the most important challenge comes from those theorists who argue for the following 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 concepts with determinate extensions.

For instance, consider 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, p. 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, p. 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 deliberatively 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, pp. 1020–1021)

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, p. 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

[23] For instance, see Schenk and McIntosh (2010).

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*.

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, p. 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, p. 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 bodies 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.[24] (2010, p. 215)

The biological constraints that Burge appeals to may well rule out some of Quine's more exotic candidates, but do they rule out either *the surfaces* or *facing surfaces* of bodies? Given our discussion of action-guiding vision above, it is very hard to see why they should.

[24] See Carey (2009, p. 99) for a brief version of this argument.

Even if we agree that successfully picking up and eating an *apple*—not just its facing surface—constitutes an action that fulfills a basic whole-animal function, that still leaves us with two equally acceptable ways of assigning extensions to those visual states that inform the relevant 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 a solid apple-shaped body with unseen (and graspable) parts, including insides. According to the alternative assignment, these representations pick out the solid apple-shaped body itself, representing it as having unseen (and graspable) parts. 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. In both cases, what we get are representations that are sensitive to the boundaries of the whole object, its cohesiveness, and its solidity, and thus in neither case are we in a better position to explain the relevant data, including the realization of whole-animal functions, such as eating, mating, and so on.[25,26]

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 is 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, pp. 126–127)

Might we rule out the thesis that *visual* reference is inscrutable by adopting the first-person perspective? Perhaps, by simply introspecting on my visual experience, I can reliably discern whether it refers to a three-dimensional body or just its facing surface.

I doubt that first-person access can serve this 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 the water. Given our assumption that there are various, equally acceptable, ways of assigning extensions to her visual state, and given our assumption that introspection is a reliable source of evidence concerning the extensions of our

[25] I take it that the same point applies to teleological theories of content (e.g. the theories defended by Dretske 1986, 1988; 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.

[26] For a detailed and persuasive reply to Burge's central argument, see DeChant & Quilty-Dunn (ms).

visual states, it seems reasonable to expect that her introspective judgment will be sensitive to whichever interpretation—among the group of equally acceptable ones—is rendered salient by her interests. And that is arguably what we find.

If I ask *S* to characterize her visual state, 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 will 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 sort 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 post-perceptual *belief*. But if that is the case, why take them to be reliable guides as to whether visual reference is inscrutable in the first place?

7 Inscrutability and object-seeing

If I'm right that visual reference is inscrutable, does it follow that it is indeterminate what 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? Rather than looking for a way to avoid it, I want to conclude by briefly suggesting that the inscrutability of object-seeing is compatible with folk intuitions.

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, p. 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).

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. 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.[27]

[27] Some theorists have argued that whether *S* sees O (over and above its facing portion) is a context-sensitive affair: e.g. see Clarke (1965), Neta (2007), and Phillips (2015). If contextualism is right then that gives us another way to explain how the inscrutability of visual reference is compatible with folk intuitions

8 Conclusion

When it comes to thought and natural language, the thesis that we cannot determinately refer to ordinary 3D objects is highly 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 distinction between thought and perception is that, while the former determinately represents objects and their facing portions, the latter admits of multiple, equally acceptable, content assignments.

Acknowledgements I would like to thank the following people for helpful input on various versions of this paper: Ryan DeChant, Uriah Kriegel, Jesse Prinz, Jake Quilty-Dunn, David Rosenthal, and three anonymous referees. I would also like to thank Laura Larocca for her help with the figures.

References

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*, 3111–3118.

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.

Burge, T. (2011). Border crossings: Perceptual and post-perceptual object representation. *Behavioral and Brain Sciences*, *34*(3), 125.

Bushnell, B., Harding, P., Kosai, Y., & Pasupathy, A. (2011). Partial occlusion modulates contour-based shape encoding in primate area V4. *Journal of Neuroscience*, *31*, 4012–4024.

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*, 329–337.

Clarke, T. (1965). Seeing surfaces and physical objects. In M. Black (Ed.), *Philosophy in America* (pp. 98–114). London: George Allen & Unwin Ltd.

DeChant, R., & Quilty-Dunn, J. (ms). Perceptual content, indeterminacy, and natural selection: Defending Quine against Burge.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.

Dretske, F. (1986). Misrepresentation. In R. Bogdan (Ed.), *Belief: Form, content and function* (pp. 17–36). New York: Oxford University Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

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*, 161–177.

Evans, G. (1975). Identity and predication. In *Collected papers*. Oxford: Clarendon Press.

Fodor, J. (1994). *The elm and the expert*. Cambridge, MA: MIT Press.

Fodor, J. (2007). Revenge of the given. In B. McLaughlin & J. Cohen (Eds.), *Contemporary debates in philosophy of mind* (pp. 105–116). Oxford: Blackwell.

Footnote 27 continued
about object-seeing. In brief, which reference-scheme (among the list of equally acceptable ones) the folk go with will simply vary according to their interests as ascribers.

Gates, G. (1996). The price of information. *Synthese*, *107*(3), 325–347.
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), 1–13.
Jackson, F. (1977). *Perception: A representative theory*. Cambridge: Cambridge University Press.
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*, 2463–2475.
Kastner, S., & Ungerleider, G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.
Kellman, P., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.
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.
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. (2006). *The visual brain in action* (2nd ed.). Oxford: Oxford University Press.
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* (2nd ed.). Cambridge, MA: MIT Press.
Nanay, B. (2010). Perception and imagination: Amodal perception as mental imagery. *Philosophical Studies*, *150*, 239–254.
Neta, R. (2007). Contextualism and a puzzle about seeing. *Philosophical Studies*, *134*(1), 53–63.
Pautz, A. (2013). Does phenomenology ground mental content? In U. Kriegel (Ed.), *Phenomenal intentionality* (pp. 194–234). Oxford: Oxford University Press.
Peacocke, C. (2014). Perception, biology, action, and knowledge. *Philosophy and Phenomenological Research*, *88*(2), 477–484.
Phillips, B. (2015). Contextualism about object-seeing. *Philosophical Studies*. doi:10.1007/s11098-015-0619-6.
Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, *80*(1/2), 127–158.
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.
Rensink, R. A., & Enns, J. T. (1998). Early completion of occluded objects. *Vision Research*, *38*, 2489–2505.
Schenk, T., & McIntosh, R. D. (2010). Do we have independent visual streams for vision and action? *Cognitive Neuroscience*, *1*, 52–62.
Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*, 259–290.
Scholl, B., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159–177.
Searle, J. (1987). Indeterminacy, empiricism, and the first person. *The Journal of Philosophy*, *84*(3), 123–146.
Spelke, E. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.
Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature*, *401*, 269–272.
Tse, P. (1999). Volume completion. *Cognitive Psychology*, *39*, 37–68.
van Lier, R., & Gerbino, W. (2014). Perceptual completions. In J. Wagemans (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), 1721–1741.
van Marle, K., & Scholl, B. (2003). Attentive tracking of objects and substances. *Psychological Science*, *14*(5), 498–504.
Wu, W. (2008). Visual attention, conceptual content, and doing it right. *Mind*, *117*(468), 1003–1033.