Contextualism about object-seeing

Ben Phillips

Springer

Springer

CrossMark

Contextualism about object-seeing

Ben Phillips[1]

Abstract When is seeing part of an object enough to qualify as seeing the object itself? For instance, is seeing a cat's tail enough to qualify as seeing the cat itself? I argue that whether a subject qualifies as seeing a given object varies with the context of the ascriber. Having made an initial case for the context-sensitivity of object-seeing, I then address the contention that it is merely a feature of the ordinary notion. I argue that the notions of object-seeing that earn their explanatory keep in both vision science and the philosophy of perception are context-sensitive as well.

Keywords Object-seeing · Contextualism · Seeing-ascriptions · Multiple-object tracking · Perceptual demonstrative thought

1 Introduction

It almost goes without saying that we have the capacity to see ordinary objects such as cats, pistachio nuts, and bowling balls—just to name a few. Folk psychology is certainly shot through with ascriptions of seeing ordinary objects, and you will be hard-pressed to find a philosopher or vision scientist who denies that we can see them. But notice that accepting this seemingly innocuous claim generates a very interesting question, for in order to see an ordinary object, one needn't see every single part of it. In order to count as seeing you, I do not have to see your liver, your heart, nor any other parts that are occluded by your facing surface. But presumably I'm able to see more than just your facing surface. I'm able to see *you*! So when is seeing part of you enough to qualify as seeing you? Suppose you are hiding behind the couch. If I can see the top of your head is that enough to qualify as seeing *you*?

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

[1] The Graduate Center, CUNY, 365 Fifth Ave, New York, NY 10016, USA

What about an ear or the very tip of your nose? In general, when does seeing part of an object suffice to count as seeing the object itself?

In what follows, I defend a contextualist answer to these questions. That is to say, I defend the view that whether *S* qualifies as seeing *O* varies according to the interests of ascribers: relative to your interests, *S* might truly be said to see *O*; whereas, relative to mine, she might truly be said to only see *O*'s facing portion. Importantly, I argue that the context-sensitivity of object-seeing is not just limited to the ordinary notion. First, I argue that explanations of phenomena that philosophers of perception have traditionally been concerned with (e.g. reference-fixing for perceptual demonstrative thoughts) invoke a context-sensitive notion of object-seeing. I then argue that which head-world relation is expressed by the notion of object-seeing that earns its keep in vision science varies according to the explanatory interests of the theorist.

I proceed as follows. In Sect. 2, I present the primary source of evidence for the context-sensitivity of the ordinary 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 Sect. 5 and Sect. 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 Initial evidence for the context-sensitivity of the ordinary notion

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.

2.1 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."

2.2 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."

2.3 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."[1]

3 A contextualist account of the ordinary 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 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.[2]

---

[1] For some similar cases, see Dretske (1969, 27) and Neta (2007).

[2] Dretske (1969, 28) seems to suggest a view along these 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

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.[3]

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 degree to which I see Venus.[4]

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?

Footnote 2 continued

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."

[3] See Kennedy (1999) for a detailed discussion of scalar analyses.

[4] 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.

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 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."[5] (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 containing 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

---

[5] See Phillips (2014) for a more detailed discussion of the notion of high integration, as well as the way in which Kriegel deploys it.

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.

3.3 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 always sufficient for seeing the object itself.

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 12:30 pm" by pointing out that it is, in fact, 12:33 pm. In reacting to your challenge I will say something like, "Well, yes, I was speaking loosely. I realize that the exact time is 12:33 pm."

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. 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 is concede that I was being hyperbolic or speaking metaphorically.[6]

3.4 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."

[6] 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.

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.

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 happens 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?

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.

4.2 Contextualism and reference-fixing for perceptual 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.[7]

---

7 See Strawson (1959) for a version of this view.

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 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*).[8,9]

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 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).

---

[8] 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.

[9] 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).

Further support for the differentiation condition comes from a consideration of the action-guiding role of vision. 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.

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$_n$ 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.

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 it, 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.[10]

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, 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.[11] 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, it is doubtful that appealing to segmentation and grouping processes will give us context-invariant notions of visual differentiation and seeing. For 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 along the horizon. Both the car and its facing surface are cohesive, bounded, and spatiotemporally 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, 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.

---

[10] Once again, I will not argue for the view that which properties are represented in perception (or thought) is a context-*invariant* affair. In brief, though, one reason for denying that which properties are represented in *perception* 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 which properties my experience represents does not vary in this way either. But that is a topic for another paper.

[11] This appears to be Burge's view of seeing (2009, 31, 2010, 456).

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.

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, 2014). 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 relations 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 ordinary

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.[12] 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. If there were a unique head-world relation that is privileged by the explanations of vision scientists then, at best, that would provide support for the ambiguity view.

5.2 The ambiguity view

How might one argue for the view that, unlike the folk notion of seeing, the one deployed 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.

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.

6.1 Individuating biological traits and individuating states of seeing

In order to motivate the view that the vision scientist's explanatory interests determine which 'seeing' relation she is invoking, I want to start by briefly

---

[12] See Putnam (1975) and Kripke (1980).

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 interest-independent way to do it.[13]

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

---

[13] See Nanay (2010, 2011, 2012, 2015) for a defense of this view.

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 relationalists (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 paper. 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 is 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 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.

6.2 Multiple-object tracking (MOT) and explanatory interests

In a series of studies, Pylyshyn (2001, 2003, 2006, 2007) has provided compelling evidence that ordinary subjects can track up to 4–5 objects (among a field of distractors). Moreover, he provides extensive evidence for the view that we do not track objects in virtue of representing their colors, sizes, shapes, or kinds. If we did track objects by visually representing these sorts of properties, we would expect performance to improve when targets and distractors are distinguishable in the relevant respects; but experiments show that this does not happen.

This has led Pylyshyn to posit a mechanism that picks out targets and maintains referential contact with them via the assignment of "visual indexes": singular elements akin to pure indexicals (e.g. "this" and "that"). Moreover, 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, 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-, location- or kind-properties right; but tracking an object does require that early vision has discriminated it from its immediate surroundings (I elaborate on this issue in more detail below).

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. 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 dismal.[14]

One way to avoid these causal indeterminacies—without relinquishing the view that visual indexes track objects without encoding their colors, shapes, and sizes—would be to abandon a purely causal theory and adopt 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

---

[14] 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.

> 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)

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.[15]

We are thus back where we were above, armed with a differentiation condition that does not determinately distinguish between ordinary 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 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 surface—as the object picked out by my visual indexing system.[16] 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 non-descript 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*: the capacity to perceive objects as having occluded parts. For instance, suppose you were confronted with the scene depicted below in Fig. 1.

Consider the claim that you saw the partially occluded face, over and above its visible portion: call the relevant relation $seeing_1$. Contrast this to the claim that you only saw the visible portion of the face: call the relevant relation $seeing_2$.

---

[15] See Clark (2006) and Dickie (2010) for similar points.

[16] It is well known that top-down attention can drive the selection of targets. See Pylyshyn and Annan (2006) for a discussion.
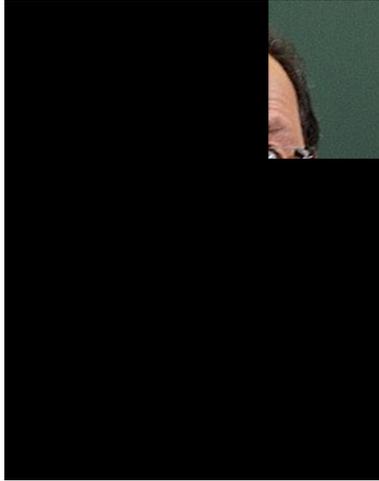
Fig. 1 Partially occluded face

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 *seeing₂* 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-shaped surface, she will naturally construe you as having seen the face itself (i.e. she will invoke the *seeing₁* relation): whether you recognized this partially occluded surface *as Hollande's face* is not a relevant concern in this explanatory context.

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 do so, we are in danger of searching for a privileged relation where there is none.

References

Block, N. (2013). The grain of vision and the grain of attention. *Thought, 1*, 170–184.

Block, N. (2014). Seeing-as in the light of vision science. *Philosophy and Phenomenological Research, 89*(3), 560–572.

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.

Campbell, J. (2002). *Reference and consciousness*. Oxford: Clarendon Press.

Clark, A. (2006). Attention and inscrutability: A commentary on John Campbell. *Reference and consciousness*, for the Pacific APA Meeting, Pasadena, California, 2004, vol. 127, pp. 167–93.

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

DeRose, K. (2009). *The case for contextualism: Knowledge, skepticism, and context* (Vol. 1). Oxford: Oxford University Press.

Dickie, I. (2010). We are acquainted with ordinary things. In R. Jeshion (Ed.), *New essays on singular thought* (pp. 213–245). 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.

Gates, G. (1996). The price of information. *Synthese, 107*(3), 325–347.

Grice, H. P. (1961). The causal theory of perception. In *Proceedings of the Aristotelian Society*, vol. 35, pp. 121–68.

Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Oxford University Press.

Jackson, F. (1977). *Perception: A representative theory*. New York: Cambridge University Press.

Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. New York: Garland.

Kriegel, U. (2009). *Subjective consciousness: A self-representational theory*. New York: Oxford University Press.

Kripke, S. (1980). *Naming and necessity*. Oxford: Basil Blackwell.

Matthen, M. (1998). Biological universals and the nature of fear. *Journal of Philosophy, 95*, 105–132.

Nanay, B. (2010). A modal theory of function. *Journal of Philosophy, 107*, 412–431.

Nanay, B. (2011). Function, modality and mental content. *Journal of Mind and Behavior, 32*, 84–87.

Nanay, B. (2012). Function attribution depends on the explanatory context. *Journal of Philosophy, 109*, 623–627.

Nanay, B. (2015). The representationalism versus relationism debate: explanatory contextualism about perception. *European Journal of Philosophy, 23*(2), 321–336.

Neta, R. (2007). Contextualism and a puzzle about seeing. *Philosophical Studies, 134*(1), 53–63.

Phillips, B. (2014). Indirect representation and the self-representational theory of consciousness. *Philosophical Studies, 167*(2), 273–290.

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science, 7*, 215–271.

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.

Pylyshyn, Z., & Annan, V. (2006). Dynamics of target selection in multiple object tracking (MOT). *Spatial Vision, 19*, 485–504.

Siegel, S. (2006). How does phenomenology constrain object-seeing? *Australasian Journal of Philosophy, 84*(3), 429–441.

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.