teorema
Vol. XXVIII/1, 2009, pp. 5-20
[BIBLID 0210-1602 (2009) 28:1; pp. 5-20]

Visual Transparency

Jeff Engelhardt

RESUMEN

De acuerdo con Roderick Chisholm, una característica distintiva de los fenómenos mentales consiste en que se relacionan con sus objetos "bajo un aspecto". Lois Lane admira como "Superman", pero no como "Carl Kent", al objeto que es tanto Superman como Carl Kent. En este artículo se argumenta que la atención visual no exhibe tal característica. Esto sugiere que la atención es un fenómeno mental excepcional y que entenderla podría proporcionar ciertas intuiciones sobre las relaciones entre la mente y el cuerpo o entre la mente y el mundo.

PALABRAS CLAVE: *filosofía de la ciencia cognitiva, percepción, reducción, relación mente-mundo, causación mental.*

ABSTRACT

According to Roderick Chisholm, one distinctive characteristic of mental phenomena is that they relate to their objects "under an aspect"': Lois Lane admires the one object that is both Superman and Clark Kent as "Superman" but not as "Clark Kent". This paper argues that visual attention exhibits no such characteristic. This suggests that attention is an exceptional mental phenomenon, and understanding it may provide insight into mind-body or mind-world relations.

KEYWORDS: *philosophy of cognitive science; perception; reduction; mind-world relation; mental causation.*

I. INTRODUCTION

Not long into *Death Wish V: The Face of Death*, Olivia Regent and Paul Kersey (Charles Bronson's character) have the following exchange:

> [OLIVIA] Paul, what do you see when you look at me?
> [PAUL] I see the woman I'm going to marry.

Paul's response suggests that verbs referring to vision provide a "transparent context". Given, that is, that (i) Paul sees Olivia, and (ii) Olivia is "the woman [Paul] is going to marry," it follows that (iii) Paul sees the woman he's going to marry. On the other hand, it is dubious that Paul also sees the woman who will later be killed by Freddie Flakes, even though Olivia is that woman too.[1] In this case, "...sees..." provides an "opaque context". Supposing, with Chisholm and others, that verbs capable of providing a transparent context refer to non-mental phenomena and others refer to non-physical phenomena[2], a puzzle thus arises: are visual phenomena mental?

On its face, the question is semantical. If "S sees *x*" is apt only if S's visual experience of *x* is available to S for verbal report, then "...sees..." provides an opaque context, it refers to a paradigmatic class of mental phenomena, and the answer is "yes". If "S sees *x*" entails only that *x* appeared unobstructed in S's visual field in good light, that S is a "normal" viewer, etc., then the answer is "no" since "...sees..." provides a transparent context. But a substantive question lingers: is the fact that "S sees *x*" admits of both of these interpretations suggestive? Are each of these interpretations rooted in some truth about the nature of visual processing? The following paper takes it that they are. In particular, it is argued that the mechanism for *visual attention* relates to its targets non-aspectually, i.e. independent of how they are described. Paradigmatic mental phenomena, on the other hand, relate to their objects aspectually: Olivia loves Paul, for example, under the description "hero", but not as "the man who will soon kill Freddie Flakes". Since "...attends visually to..." provides a transparent context, and granted that most visual processes relate to the world aspectually[3], it is unsurprising that verbs referring to both attention and the "higher level" processes it enables should provide both transparent and opaque contexts. Furthermore, if it is true that verbs providing transparent contexts refer to non-mental phenomena, then it follows that even though most visual phenomena are presumably mental, visual attention is non-mental.

II. Aspects and Opaque Contexts

What is an opaque context? Let us illustrate it by reference to an example. Ask yourself whether 1 and 2 *imply* 3.

1. John believes that Batman is tough.
2. Batman is Bruce Wayne.
3. John believes that Bruce Wayne is tough.

I assume you've answered yourself that 1 and 2 do not imply 3. The only difference between 1 and 3, though, is that "Bruce Wayne" is substituted for "Batman". And though substituting terms in and out of true claims seldom results in true claims, the result is surprising in this case because, as 2 says, Bruce Wayne *is* Batman. "Bruce Wayne" and "Batman" are co-referential.

This seems to be a general feature of sentences featuring "...believes...": starting with a true sentence, substituting a co-referential term into the complement clause following "believes" doesn't always generate another true sentence. In other words, the claim that a person A believes something P about an *x*, coupled with the information that *x* is *y* doesn't entail that A believes P of *y*. "...believes...", then, provides an opaque context. We cannot *see* the truth of 3 *through* the truth of 1 and 2. This characteristic seems to be common to verbs referring to mental phenomena.[4]

Verbs that refer to physical phenomena, however, exhibit no such general characteristic. Ask yourself whether 4 and 5 imply 6.

4. John punches Batman.
5. Batman is Bruce Wayne.
6. John punches Bruce Wayne.

Here, 4 and 5 *do* imply 6. We *can* see 6 through 4 and 5. "...punches...", then, provides a transparent context. It is sometimes thought that this difference in the entailments licensed by sentences featuring certain verbs suggests a difference in the phenomena to which the verbs refer. [See, for example, Chisholm (1957), pp.168-185] In particular, it is claimed that verbs that provide an opaque context refer to non-physical phenomena while verbs providing a transparent context refer to non-mental phenomena. So, if you want to know whether a verb refers to a non-mental or non-physical phenomenon, you can apply this test:

(a) Select a sentence, S, in which the verb to be tested takes a direct object or complement clause, P.
(b) Select a term, P*, that is co-referential with P.
(c) Substitute P* for P in S to create a new sentence, S*.
(d) Determine whether the conjunction of S and the sentence "P and P* are co-referential" entails S*.
(e) If "no", the verb refers to a non-physical phenomenon.
(f) If "yes", the verb refers to a non-mental phenomenon.

If Chisholm is right about transparent and opaque contexts, this test will tell us whether the referent of a verb is non-mental or non-physical. Supposing that it does work, though, it is reasonable to ask why it works. Why do verbs referring to non-physical phenomena provide an opaque context? When it comes to the non-physical phenomena *we're* interested in, mental phenomena, the reason often cited is that they relate to those things they're about "under an aspect" or "under a description". The belief expressed in "John believes that Batman is tough," for example, relates to the one object that is both Batman and Bruce Wayne under the aspect or description "Batman". It is assumed, reasonably, that facts about aspects or descriptions of an object need not imply anything about one another: John, like much of Gotham City, may not know that Batman and Bruce Wayne are one and the same. Hence, John may also think that Bruce Wayne is not tough in the least without harboring inconsistent beliefs. Verbs referring to mental phenomena provide opaque contexts, then, because mental phenomena relate to objects under aspects, and aspects of an object do not carry information about one another, even if they are aspects of the same object.

III. "...SEES..."

Now, let us apply our test to "...sees...". If it provides an opaque context, then vision relates to things seen aspectually, and visual processes are, plausibly, mental. 7 is our S; "Batman" and "Bruce Wayne" are our P and P*, respectively; and, 9 is S*. Now we need to determine whether 7 and 8 imply 9.

7. John sees Batman.
8. Bruce Wayne is Batman.
9. John sees Bruce Wayne.

It is tempting to say that seeing Batman *does* imply seeing Bruce Wayne, since there's only one thing to be seen in the first place. On the other hand, if John doesn't know that Batman is Bruce Wayne, he would deny 9. So it's also tempting to say that 7 and 8 do not imply 9, and especially if you think that "S sees *x*" entails "S is able to report seeing *x*".

It seems that our test has failed: "...sees..." provides ambiguous results. We need some other means for determining whether vision relates to its objects aspectually, and whether visual processes are mental phenomena. I propose two steps for answering these questions. First, we should replace seeing with a more specific visual process about which there are empirical results available for consultation: visual attention. Second, we should consult these empirical results rather than our linguistic intuitions, asking them, in particular, whether visual attention relates to its objects aspectually. We will see that the mechanism for visual attention does not relate to its targets under an aspect, and so the accurate use of "...attends visually to..." provides a transparent context. Hence, it will be shown that "...sees..." provides a transparent context at least insofar as it refers to visual attending, and that visual attention is plausibly a non-mental phenomenon.[5]

Before we begin, though, notice that if perceptual processes operate primarily over aspectual phenomena, then they *demand* the availability of non-aspectual phenomena. Perceptual processes are supposed, minimally, to provide an agent with information about her local environment. Much of this environment is physical, and thus non-aspectual. If the information resulting from perceptual processes *is* aspectual, then it differs from many of the objects from which it came. Presumably, then, some process "extracts" aspectual information from *something non-aspectual.* The "something non-aspectual" may be the object itself or it may be a vision-internal, non-aspectual proxy for visual targets. This paper takes no position on the matter. It claims, rather, only that a) the myriad visual processes include at least one that does not relate to its objects under an aspect, and that b) visual attention is such a process. My argument, in brief, is that the empirical data concerning the operations of visual attention suggest that it can relate to its targets independent of their aspects. As such, if a process *must* relate to its objects under an aspect in order for verbs referring to it to provide an opaque context, then "...attends visually to..." provides no such thing.

IV. Visual Attention

First, what is visual attention? Phenomenologically speaking, it is familiar enough. When you're walking down the street and a speeding car bursts into your field of view such that you can't help but follow it with your gaze, the car has "grabbed your attention", your *visual* attention. As far as the going theories on visual attention are concerned, visual attention has two roles: selection and tracking. Selection occurs when something to which you're not attending becomes something to which you do attend: when the speeding car `catches your eye'. Tracking ensues as the target of your attention moves across your visual field. In most contemporary theories of vision, attention is supposed to be a prerequisite for encoding visual information. [Pylyshyn (2003), pp. 88-89, 159-200; Posner (1980), p. 4]

For expository purposes, I will delineate the view that visual attention is not aspectual in contrast to the strongly intuitive view that vision attends to objects according to their locations, i.e. under their locative aspect. The former I shall call Visual Index Theory or simply VIT, after the mechanisms Zenon Pylyshyn has proposed as the visual executers of a "mind-world connection". [Pylyshyn (2001), pp. 183-195, (2008), pp. 1-65] According to VIT, an object in a visual field is *selected* when a connection is established between the object and an "index" in the visual system. Tracking ensues as long as this connection remains intact. Although these indexes *can* relate to their targets under an aspect, it is not necessary that they do so. [See Pylyshyn (2003), pp. 215, 218-23] Rather, a sufficient index-to-target connection need only be a causal connection, *no matter what information this connection carries.*[6]

We can illustrate this theory with a simple analogy. Imagine the shadow cast by a sundial as a visual index, the sundial being the analogue of the visual system. When it's dark, there is no shadow, and so nothing selected. When it's sunny, though, the sundial casts a shadow; by its shadow, the sundial selects an object, the sun, and as the sun moves across the sky, the shadow will move along with it, as if tracking it. I think it is anodyne to say that the sundial relates to the sun non-aspectually:

10. The sundial's shadow tracks the sun across the sky.
11. The sun is the largest body in our solar system.
12. The sundial's shadow tracks the largest body in our solar system across the sky.

I hope I'm not alone in thinking it plausible that 10 and 11 entail 12. If they do, this suggests that "...tracks..." provides a transparent context. If visual attention is similar in this respect, it relates to its targets non-aspectually, as VIT proposes.

I shall call the view that vision tracks under locative aspects the location-based theory or LBT. According to this view, visual targets are selected and tracked by a mechanism that operates over representations of their locations. Supposing that these representations must exploit information from the targets, this means that the minimally sufficient mechanism-to-target connection must carry information about an attentive target's location. Imagine looking through an automated telescope that registers on an internal map the location at which it is pointed at any time. You can't command it to select the moon; you must give it the proper coordinates. Similarly, when a `shooting star' suddenly enters its view, it doesn't just swivel and select it. Rather, the disturbance behaves just like a command: it incites the telescope to encode the location of disturbance, then turn and "attend" to that location. As the star moves across the sky, the telescope tracks it by continually encoding its new locations. Both the mechanism proposed by LBT and this telescope select and track their targets "under" their locations in that they fail unless they represent their targets' locations. A verb referring to either of these, then, should provide an opaque context. Consider:

13. The telescope selects location *l*.
14. The object at location *l* is a shooting star.
15. The telescope selects a shooting star.

LBT says selective attention and tracking are similar.[7] If so, visual attention relates to its targets aspectually, and it is a mental phenomenon.

V. An Argument for VIT

Here's where we stand. VIT is true if and only if "...attends visually to..." provides a transparent context; LBT is true just in case selection and tracking relate to their targets under their locative aspects. Now we can make this dispute more concrete and bring empirical data to bear. I take it that if visual attention relates to its targets under their locations, then it must be that the algorithm that best models its operations *must* refer to the locations of its targets.[8] If the best model of selection and tracking need not refer to locations or any other aspects of its targets, on the other hand, then visual attention is not an aspectual relation. Furthermore, I take it that we determine which algorithm is the best model for visual attention by consulting the experimental evidence that is currently available. If this evidence suggests that subjects can attend to visual targets without exploiting information about the locations of those targets, then the algorithm modeling visual attention need not refer to locations. That is, this evidence suggests that LBT is false. And the same, *mutatis mutandis*, goes for any other aspects that targets of visual attention might have. Hence, if the evidence suggests that subjects can attend to visual targets without exploiting *any* information about those targets, then the algorithm modeling visual attention need not refer to any aspects. That is, VIT is true and "...attends visually to..." provides a transparent context. I will argue that the available evidence suggests exactly this.

Presently, in V.1, I'll marshal evidence that subjects can attend to visual targets and discriminate them from "distractors" when the targets' only distinguishing aspects are their locations. It follows that the algorithm modeling visual attention can yield success without referring to aspects other than locations. Thus, the only aspect that is plausibly essential to visual attention is location; and if LBT is true, it must be possible for subjects to select and track visual targets by virtue of locative information *alone*. In V.2, I'll show that neither of the two most plausible selection and tracking algorithms mentioning only location(s) — that is, neither of the two most plausible tracking algorithms respecting the constraint argued for in V.1— can account for recent experimental results. It follows that the selection and tracking algorithm does not exploit information about any of its targets' aspects, as VIT states.

V.1

In a number of "multiple object tracking" (henceforth MOT) experiments, it has been shown convincingly that we can track up to five randomly and independently moving targets in a visual field with five visually indistinguishable "distractors" at a success rate of around 85%. To account for this capability, both VIT and LBT now affirm that we in fact have multiple tracking mechanisms operating in parallel—for we can't explain our successes in MOT tasks with serial processing of a single mechanism.

In the basic task, subjects sit before computer screens displaying ten visually indistinct objects against a uniform background, often blue circles in a black field. They're instructed to track a subset of the objects (the "target subset") until the end of the trial, say from 10 to 20 seconds. As the trial begins, the members of the target subset briefly flash off and on to grab the subjects' attention, and then all the objects in the field begin to move independently along random paths. At the conclusion of the trial, subjects use a mouse pointer to indicate which objects flashed. [See, for example, Pylyshyn and Storm (1988); Cavanagh (1999); Pylyshyn (2008), p. 24 has an extensive list of MOT results from dozens of experimenters.]

Since the targets and distractors are indistinguishable with respect to their visible aspects except location, trajectory, and velocity, it cannot be that we track the targets by dint of information about any of their other distinctive aspects. The tracking algorithm, then, need not refer to any of these aspects, and visual attention need not relate to its targets under any of them.

So what about velocity and trajectory? There are good reasons to think that our tracking mechanisms do not exploit trajectory or velocity information in order to "predict" and track targets. Indeed, all available evidence indicates that they do not. In an MOT variation, Pylyshyn and Keane (2006) showed that we track through disappearance at the usual rate *only when the objects reappear at the same place they disappeared.* The trials in this study were similar to the other MOT trials, except that at some point in each trial, all objects on the screen disappeared for a fixed duration. When the objects reappeared, there were two conditions on the objects' reappearances; call them *the displacement condition* and *the stable condition*. In the displacement condition, objects reappeared at the location they would have been had they maintained their trajectories and velocities throughout the disappearance. In the stable condition, objects reappeared at the same locations from which they disappeared. It was found that subjects did reliably better in the stable condition (over 90%) than in the displacement condition (from 65-85%). Were our tracking mechanisms exploiting a prediction strategy, we would have seen the reverse. With this and similar results from other MOT permutations, the experimenters conclude that mechanisms do not exploit a predictive strategy in tracking through disappearance [Keane and Pylyshyn (2006), pp. 353, 357-8, 362].

We succeed at simultaneously tracking multiple targets that are indistinguishable from distractors. This shows that we can track objects without exploiting their distinctive perceptual aspects other than location, velocity, and trajectory. Furthermore, we have seen results indicating that tracking mechanisms do not "predict" a disappeared target's reappearance; in the absence of countervailing evidence, there is no reason to say that tracking mechanisms exploit information regarding a target's velocity or trajectory. We should conclude, then, that either VIT is true or the visual attention algorithm refers to locative aspects alone.

V.2

Turning the tables a bit, I contend that our success in tracking through disappearance lends credence to VIT and poses a problem for LBT. Although Keane and Pylyshyn (2006) showed that subjects tracked significantly worse in the displacement condition, they also showed that there was a significant percentage of success in this condition—around 85% for a 150 ms disappearance and 65% for 450 ms. [Keane and Pylyshyn (2006), p. 353)] Similarly, it has been shown that tracking performance is hardly diminished when targets and distractors overlap one another during the trials. [Viswanathan and Mingolla (2002), pp. 1415, 1418-25].[9] No advocate of LBT has, to my knowledge, addressed this data and shown that an algorithm using only locative information could model it. One plausible approach would insert into the algorithm a strategy to take the object *nearest* to the location of disappearance as the target. I'll argue that no such strategy can yield tracking results that fit the data, and so, *prima facie*, LBT cannot model it.[10]

V.2.1 *The Spotlight Strategy*

We might suggest that the tracking mechanism selects the object nearest the location of disappearance by selecting the location of disappearance, "spreading" attention outward—perhaps to a limit—and then selecting as the target the first object it comes across. It would be as if one trains a spotlight on the target while tracking and then, upon disappearance, enlarges the spotlight's area until the target reappears, and then re-centers on it. Seizing on this metaphor, I'll call this approach the "spotlight strategy". The problem is that the spotlight strategy predicts tracking at chance in overlapping-object tasks; since we track at normal rates through overlap, this strategy requires amendment or rejection.

In figure 1 (below), a target and a distractor overlap in frame 4. Clearly, no spotlight strategy acting alone will suffice to discriminate the target and the distractor from 4 to 5. At 4, the mechanism encodes the location of the target, but this is also the location of the distractor. If the distractor and target are moving at the same speed, then the mechanism will be at chance in selecting the target over the distractor, since both will be equidistant from the location encoded in frame 4. If the distractor is slower, the mechanism will select *it*. This result holds no matter how precisely or how frequently the mechanism executing the spotlight strategy encodes locations.

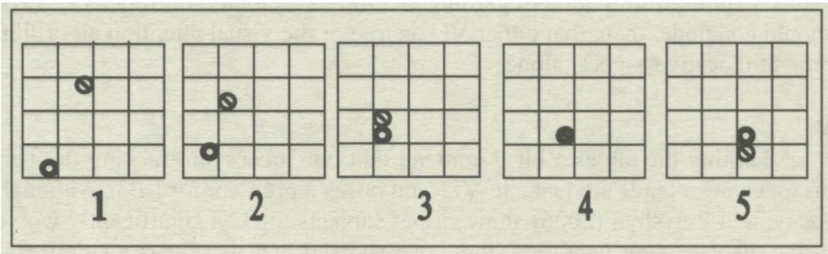

FIGURE 1: The donut represents a target, and the crossed donut represents a distractor, although the two are, of course, visually indistinguishable in the experiment. The frames represent a chronological progression starting with frame 1. Remember that the objects move randomly.

Notice that the proponent of an LBT algorithm faces a dilemma here. On the one hand, the mechanism must be so "insensitive" to locations as to abstract away from the distinct locations of its targets' time-slices. On the other, it must attend so closely to its targets' locations as to discriminate them from nearby distractors. If the mechanism trains a "wide view" on the target, it is rendered insensitive to distances between the objects: both target and distractor will appear in the view during overlap, and so the mechanism will be at chance in selecting its target at its next update. If the mechanism casts a "narrow view", even the target's slight movements will overflow its borders. Any time a distractor is near enough to touch the target, it will also border the view and so the mechanism will again be at chance between the two.

V.2.2 *The Informed Spotlight Strategy*

The spotlight strategy fails to reproduce our success in overlap tasks, but it would fare better if the mechanism for visual attention could be continually apprised of distractors' locations. For, the problem plaguing the basic spotlight strategy is that the mechanism has no means for deciding between two objects that are both within its view. If it could be informed of a distractor not far below it, the mechanism could "decide" between two objects in its view by "preferring" the one toward the top. That is, this strategy would adjust the preferences of the basic strategy so as to prefer the object farther from the last reported location of the nearest distractor. In figure 1, a mechanism making this adjustment would be more likely to choose the target. Call this the informed spotlight strategy. I'll show now that just as with the spotlight strategy, the claim that our mechanisms execute the informed spotlight strategy entails that we'll track at chance in complete overlap. Consider situations like those depicted in figure 2.[11]

When the distractor nearest a target approaches it into a complete overlap, as in 2A, or overtakes it in a complete overlap, as in 2B, a tracking mechanism executing the informed spotlight strategy would either select the wrong object when updating from a stored location or be at chance, depending on how precisely it encodes locations and how frequently it updates its encoded locations. In figure 2, the mechanisms will be at chance when forced to select in frame 3, but things might be even worse. If the mechanism does not update exactly when the objects completely overlap, it will have a preference for the object farther from where the distractor was prior to overlap. Since the distractor comes out the other side of the target, this object will always be the distractor. So if the tracking mechanism doesn't update during the moment of occlusion, this model predicts perfect failure.

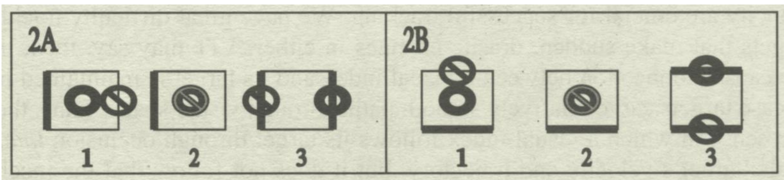

FIGURE 2: The doughnut represents a target and the crossed circle a distractor. In 2A, the objects are traveling toward one another in the first frame and away from each other in the third. In 2B, both objects travel in the same direction, but the distractor travels faster. By frame 3, the distractor has overtaken the target. At least one side of a frame is thickened in each case to indicate the direction of the mechanism's preference. Notice that frames A-3 and B-3 have two thickened sides to indicate that the respective mechanisms have no preference that distinguishes the target and distractor.

I conclude that LBT does not model the mechanism for visual attention.

VI. Objections

Objection. The results reported here depend crucially on experimental trials in which a target and a distractor occupy one and the same location in the visual field. Since subjects nonetheless reliably distinguish targets and distractors, it seems that locative information is not necessary for successful selection and tracking. Yet, there is something amiss about this interpretation. Namely, it requires us to assume that there is a fact of the matter about which visual object is the target and which the distractor during occlusion, *even though the "two" share all of their features.* The interpretation provided here assumes that we visually individuate objects that are visually indistinguishable. Yet, such an extreme view is unnecessary: we can discriminate target and distractor by the trajectories and velocities they exhibit going into the occlusion. Thus, rather than concluding that visual attention needs no information for success, we can conclude that it demands *either* locative information *or* velocity and trajectory information. This aspectual approach accommodates the experimental data without postulating a "sub-personal" mechanism the intelligence of which far outstrips the subtlety of our conscious-level faculties.

Reply. It is true that an overlapping target and distractor are most easily distinguished by their trajectories and velocities, but we must distinguish between the information *available* to the tracking mechanism and the information that the mechanism *is in fact exploiting*. The results mentioned in section V.1 suggest that the tracking mechanism does not exploit trajectory and velocity information. All the same, a little reflection reveals that trajectory and velocity are crucial for successful tracking. We have great difficulty tracking targets that make sudden, drastic changes in either. VIT may say, then, that the causal connection between a visual index and its target is maintained just in case targets carve relatively smooth paths through visual space. Thus, there is a sense in which a visual index follows its target through occlusion *thanks to* the target's velocity and trajectory, but it does not follow that the mechanism is exploiting such information about its target. Rather, it need not be making use of this information at all. And, again, in light of Keane and Pylyshyn (2006), this interpretation better fits the data. Our conclusion stands.

Objection. The thesis that opacity marks mentality is dubious. On the one hand, many non-mental terms provide opaque contexts. [Cf. FN 4] On the other, many sentences featuring terms referring to mental phenomena are transparent: "S is intelligent"; "S is thinking"; and so on. Opacity is neither necessary nor sufficient to mark terms referring to mental phenomena, and so too, *mutatis mutandis*, for transparency and terms referring to physical phenomena. The distinction doesn't deserve our epistemic regard, and it surely shouldn't motivate any claims about which phenomena are and are not mental.

Reply. The claim here does not turn on opacity being necessary or sufficient to mark mentality. We endorse neither:

> [**necessary**] If a verb does not provide an opaque context, then it does not refer to a mental phenomenon.
> [**sufficient**] If a verb provides an opaque context, then it refers to a mental phenomenon.

Our *tentative* commitment, rather, is to the much weaker claim that opacity is a mark of verbs (i) taking a direct object or complement and (ii) referring to *paradigmatic* mental phenomena and, by extension, that relating to objects under an aspect is *the norm* for mental phenomena. While it takes only a single impenetrable case to defeat either of the above claims, refutation of *these claims* demands a flood of cases. There is nothing to indicate that the flood is on its way, though, and indeed, this claim is fairly anodyne amongst contemporary philosophers of mind.[12] This is inadequate to show that the phenomena of visual attention are non-mental, of course, but it is adequate to show that they are unusual phenomena of special interest to those interested in psycho-physical or mind-world relations.

VII. Conclusion

The available evidence suggests that the algorithm that best models selection and tracking need not refer to any of its targets' aspects. According to VIT, this is because visual attention relates to its object "as an individual", via either a non-aspectual representation or via some relation that doesn't crucially involve information processing [Pylyshyn (2008), pp. 1-65)]. Hence, "...attends visually to..." provides a transparent context. So it may be true that Paul doesn't *see* Freddie Flakes's next victim when he looks at Olivia, but as long as he attends to her, he attends to Freddie's next victim as well. Visual processing relates to its objects transparently in attending to them. Additionally, if it is true that verbs referring to paradigmatic non-mental phenomena provide transparent contexts, then the phenomena involving visual attention are non-mental.

Department of Philosophy
Georgetown University
Washington, DC 20057 USA
E-mail: jme37@georgetown.edu

Notes

[1] Compare the example in Chisholm (1957), p. 171: "Most of us knew in 1944 that Eisenhower was the one in command; but although he was identical with the man who was to succeed Truman, it is not true that we knew in 1944 that the man who was to succeed Truman was the one in command."

[2] The classic formulation of this claim is Chisholm (1957), pp.168-185)

[3] If you're incredulous, see, most notably, Marr (1982), pp. 1- 21, but also Pylyshyn (2003), ch. 2.

[4] It is not, however, a characteristic of verbs referring to mental phenomena exclusively. Modal verbs, for example, may exhibit it as well. For example, 1 and 2 do not entail 3 in the following:

1. It is necessary that Bruce Wayne is Bruce Wayne.
2. Bruce Wayne is Batman.
3. It is necessary that Bruce Wayne is Batman.

[5] A provocative note: If it is true that verbs providing transparent contexts refer to non-mental phenomena, then it follows that visual attention isn't a mental phenomenon. If visual attention nonetheless *enables* visual processing over aspectual phenomena, which would presumably be mental phenomena, it seems that visual attention is a plausible candidate for the locus of the "mind-world connection". See Pylyshyn (2001) for discussion of this claim in a slightly different context.

[6] *Nota Bene*: Since no information is required, it should not be surprising that this process is not always available to conscious access and report.

[7] LBT, however, doesn't require that the locations be encoded in objective coordinates. A slightly more advanced model, which might seem more like *our* system, could respond to positions relative to other objects, like "below and to the right of Hesperus". This would still require that a *location* be encoded in order for visual selection to occur.

[8] And, *mutatis mutandis*, for any aspectual view of visual attention. If you think visual objects are selected and tracked under their colors, then references to their colors should be ineliminable from the algorithm that best models selection and tracking.

[9] The authors report, "the results show that, although the tracking does become more difficult when element boundaries are allowed to intersect, it does not become impossible, even in the purely two-dimensional case" [Viswanathan and Mingolla (2002), p.1418].

[10] It is helpful to note the information available to LBT in handling overlapping object tasks. It may seem that an LBT algorithm could exploit depth cues exhibited by targets and/or distractors in its tracking during overlap. If the mechanism could discern which object overlaps which, it could maintain unique locative descriptions of each, and thus overlap would pose no trouble at all for tracking an object by its location. This intuition seems bolstered by the report from Viswanathan and Mingolla (2002) that tracking through overlap is *more successful* when objects on the screen exhibit occlusion cues. For it seems that the information carried by the occlusion cues is information about each object's location, and that the mechanism is encoding this information to aid in tracking. But we must not confuse depth *cues* with actual depth. Remember that LBT is true only if the tracking algorithm mentions only locations. If the targets and distractors *actually* moved in depth, then the objects would exhibit distinct locations and the tracking mechanism could represent them as such. Depth *cues*, however, are *marks on the objects* and not locations. As such, a mechanism representing only locations does not exploit the information they carry, and the evidence that they aid in tracking tasks is irrelevant to our discussion insofar as we are trying to discern the *minimum* of information the visual attention mechanism requires for success.

[11] This strategy requires not only the bold claim that tracking mechanisms "communicate" their encoded locations to one another, but also that the locations of

distractors are somehow encoded and communicated to the tracking mechanisms. How can the tracking mechanisms obtain *that* information? The matter is complex, but there is evidence that we're *inhibiting* distractors during MOT trials. [Flombaum, Scholl, and Pylyshyn (under review)] When we inhibit something, it takes longer for us to switch attention to it. For example, if you're scanning a scene for a squirrel and you look in a particular tree and then move your attention elsewhere, it'll take you longer to notice something in that tree than it would for you to notice something in a tree you hadn't yet attended to during the search. (This is called "inhibition of return". It is described in Klein (2000).) If distractors are inhibited as they move, then it seems that they're being tracked to some degree, although not by a selection and tracking mechanism. (That distractors are tracked in this way doesn't mitigate the problem sketched for the spotlight strategy operating in isolation. If both the target and the distractor are tracked by their locations, the fact that they'll be equidistant from the most recently encoded location will still result in tracking at chance.) Even though we are unable to report the information used for inhibition to an experimenter, it may still be the case that the mechanisms inhibiting distractors can report their locations to the mechanisms tracking targets.

[12] But see Soames (2002) for the dissenting view.

REFERENCES

CAVANAGH, P. (1999), "Attention: Exporting Vision to the Mind", in Taddei-Ferretti, C. and Musio, C. (eds.), *Neuronal Basis and Psychological Aspects of Consciousness*, Singapore, World Scientific, pp. 129-43.

CHISHOLM, R. (1957), *Perceiving: A Philosophical Study*, Ithaca, Cornell University Press.

ERIKSEN, C.W. and ST. JAMES, J.D. (1986), "Visual Attention Within and Around the Field of Focal Attention: A zoom Lens Model", *Perception & Psychophysics*, vol. XL, pp. 225-240.

FODOR, J.A. (2009). "So What's So Good About Pylyshyn?", in Dedrick, D. and Trick, L. (eds.), *Computation, Cognition, and Pylyshyn*, Cambridge, MIT Press, *Introduction.*

FLOMBAUM, J.I., SCHOLL, B.J., and PYLYSHYN, Z.W. (under review), "Attentional Resources in Visual Tracking Through Occlusion: The High-Beams Effect".

KEANE, B.P. and PYLYSHYN, Z.W. (2006), "Is Motion Extrapolation Employed in Multiple Object Tracking? Tracking as a Low-Level, Non-Predictive Function", *Cognitive Psychology*, vol. LII, pp. 346-368.

KLEIN, R. (2000), "Inhibition of Return", *Trends in Cognitive Science*, vol. IV(4), pp. 138-147.

MARR, D. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco, W.H. Freeman.

POSNER, M. I. (1980), "Orienting of Attention", *Quarterly Journal of Experimental Psychology*, vol. XXXII, pp. 3-25.

PYLYSHYN, Z.W. (2001), "Connecting Vision and the World: Tracking the Missing Link", In Branquinho, J. (Ed.), *The Foundations of Cognitive Science*, Oxford, Clarendon Press, pp. 183-195.

— (2003), *Seeing and Visualizing: It's Not What You Think*, Cambridge, MIT Press.
— (2007). *Things and Places: How the Mind Connects With the World*, Cambridge, MIT Press (Jean Nicod Lecture Series).
PYLYSHYN, Z.W. and STORM, R.W. (1988), "Tracking Multiple Independent Targets: Evidence for a Parallel Tracking Mechanism", *Spatial Vision*, vol. III, pp. 1-19.
SCHOLL, B.J. (2001), "Objects and Attention: The State of the Art', *Cognition*, vol. LXXX, pp. 1-46.
SOAMES, S. (2002), *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*, Oxford, Oxford University Press.
TREISMAN, A. (1998), "Feature Binding, Attention, and Object Perception", *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. CCCLIII, pp. 1295-1306.
TREISMAN, A., and SCHMIDT, H. (1982), "Illusory Conjunctions in the Perception of Objects", *Cognitive Psychology*, vol. XIV, pp. 107-141.
VISWANATHAN, L. and MINGOLLA, E. (2002), "Dynamics of Attention in Depth: Evidence from Multi-Element Tracking", *Perception*, vol. XXXI, pp. 1415-1437.