UC San Diego
UC San Diego Previously Published Works

Title
Objects, places, and perception

Permalink
https://escholarship.org/uc/item/4qs9w5s0

Journal
Philosophical Psychology, 17(4)

ISSN
0951-5089

Author
Cohen, Jonathan

Publication Date
2004-12-01

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Objects, Places, and Perception

Jonathan Cohen*

The familiar material objects may not be all that is real, but they are admirable examples ([Quine, 1960], 3).

Abstract

In [Clark, 2000], Austen Clark argues convincingly that a widespread view of perception as a complicated kind of feature-extraction is incomplete. He argues that perception has another crucial representational ingredient: it must also involve the representation of "sensory individuals" that exemplify sensorily extracted features. Moreover, he contends, the best way of understanding sensory individuals takes them to be places in space surrounding the perceiver.

In this paper, I'll agree with Clark's case for sensory individuals (§1). However, I shall argue against his view of sensory individuals as places (§2). Instead, I'll propose and defend an alternative account of sensory individuals that construes the latter as (visual) objects (§§3–5).

On one venerable view, perception amounts to something like a kind of feature extraction. According to this view, perception extracts and represents distal features of the world so that the organism can think and act in environmentally appropriate ways. For example, this view has it that perception registers the presence of *redness* in the environment on some occasions, and the presence of *triangularity* in the environment on others, and then passes this information to other systems so that the perceiving organism does what is appropriate for the current environment — sticking around to eat the things it should eat or fleeing from the things that would eat it, as the case may be. In [Clark, 2000], Austen Clark argues convincingly that this view of perception as a complicated feature-extractor is incomplete. He argues that perception has another crucial representational ingredient: it must also involve the attribution of the extracted sensory features to individuals — in effect, that perception must have a kind of subject-predicate structure so that it can represent that sensory features hold of certain individuals. This conclusion, which I shall call the thesis of sensory individuals, or SI, seems to me an extremely deep point about perception, and one that deserves substantial attention.

In addition, Clark has a related proposal about just how we should understand the sensory individuals the need for which he has urged. Namely, he

*Department of Philosophy, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0119, joncohen@aardvark.ucsd.edu

suggests, sensory individuals are *places*, i.e., places in space surrounding the perceiver. Call this the thesis of sensory places, or SP. Although I am extremely sympathetic to Clark's arguments for SI, I am not persuaded by SP. Indeed, I believe there is a variety of evidence that tells against SP. Moreover, it strikes me that there is a more attractive alternative proposal that I'll call SO; namely, we can suppose that the individuals to which perception attributes sensory qualities are (most often) *objects*.

In this paper I'll start by briefly reviewing Clark's arguments for SI (§1). Next, I'll consider, and take issue with, his second proposal, SP (§2). Finally, I'll lay out and defend SO — the alternative view that construes the individuals of perception as objects of a particular kind (§§3–5).

1 Sensory Individuals

As noted, many have taken perception to be a process of extracting and representing distal features, such as *redness* and *triangularity*. Clark's point is not to deny the importance of the questions about these features and how they are represented; indeed, he has articulated and defended an important account of sensory features in [Clark, 1993], and he relies on that account in [Clark, 2000] as well. Clark's contention is not that such work on sensory features is misguided, but that it is incomplete. In particular, his view is that this work needs to be supplemented by a theory of sensory individuals.

Clark motivates the need for sensory individuals through consideration of a range of problems that have gone under the name 'the binding problem'. The most important of these for Clark (henceforth, 'the Binding Problem') concerns combinations of sensory features.[1] Subjects can pick out the lone red triangle from a visual display containing red squares and green triangles. To do this, the target must have been identified on the basis of *both* its color and its shape; either, alone, would fail to distinguish the target from the distractors. That is, it is insufficient for successfully completing the task for a subject to represent the color of each element in the visual display alone (that would distinguish the target from all of the green elements, but would not distinguish the target from all of the red squares). Similarly, it won't suffice merely to represent the shape of each element alone (that would distinguish the target from all of the squares, but would not distinguish the target from all of the green triangles). What would distinguish the target from the other elements in the display would be a representation of the color *and* shape of each of them. But this means that the representation of each element's color must be bound to the representation of that element's shape: the subject must represent that there is *one thing* and that *it* is red *and* triangular. Nothing less, and in particular, representing merely that there is something red and something triangular, won't succeed in distinguishing the target from the distractors. Thus, in order to solve the Binding Problem, perceptual systems must, in effect, attribute multiple sensory predicates (shape, color, etc.) to one and the same entity. And this, in turn, means that perception must have a way of singling out entities that is independent of, and not reducible

to, its representation of features.

Of course, it is uncontroversial that the subject's visual system must represent the instantiation of *redness* and the instantiation of *triangularity*. Clark is not questioning whether these properties are represented. Rather, he is asking whether the representation of these properties, by itself, suffices. His reason for answering this question negatively is that, as we have seen, completing the experimental task requires representing that the two properties are co-instantiated, and that a mere feature extractor cannot deliver this information. We can make this point clear by adverting to familiar logical notation (needless to say, neither Clark nor I claims that perceptual representation has anything like this format). A positive result from a *redness* extractor asserts this:

(1) something is red.

Similarly, a positive result from a *triangularity* extractor asserts that

(2) something is triangular.

But what the perceptual system needs to represent in order to complete the task is

(C) Some one thing is both red and triangular.

However, just as (C) does not follow from (1) and (2), the positive results from our two feature extractors fail to give the perceptual system what it needs to represent in order to complete the experimental task. Thus, Clark concludes, feature extraction, by itself, is insufficient to do what the perceptual system needs to do (and routinely does).

We have said that, for Clark, the solution to the Binding Problem we've run up against lies in the introduction of representations for entities along the lines of what singular terms do in natural language. How will this addition help? Suppose that we combine the representational power of feature extractors with devices for representing individuals, such as the names 'a' and 'b'. Then we can express what the *redness* extractor tells us not as a general quantificational statement, but as a simple singular predication:

(1′) a is red.

Similarly, we can express what the *triangularity* extractor tells us as

(2′) b is triangular.

This is still insufficient to yield (C) by itself, but now that we have names for individuals as an expressive resource, we are in a position to get what we need by the addition of one more simple (two-place) predication:

(3) $a = b$.

By allowing that perception can represent something like (1′), (2′), and (3), we have finally endowed the perceptual system with sufficient expressive power: just as (C) does in fact follow from premises (1′), (2′) and (3), the perceptual system that can make singular predications to individuals and represent identity claims involving names for those individuals has what it needs to distinguish the red triangles from the distractors, as needed.[2] [3]

2 Against Places

Suppose we join Clark in endorsing SI — in holding that, in order to solve the Binding Problem, visual perception must deliver not simply the information that there is local redness and triangularity, but that some particular entity is red and some particular entity (possibly the same entity) is triangular. A natural next question is this: what are these particular entities to which vision attributes features? Clark's answer to this question, which I've labeled SP, is that sensory individuals are places in the three-dimensional space around the perceiver.[4]

Now, Clark's development of SP in [Clark, 2000] is tied very closely to his argument for SI. So quick is his transition between these theses that it is possible to fail to notice that there really are two separate moves being made. One theme of this paper is that Clark has run together the two questions about the subject-predicate structure of perception and the identification of sensory individuals with places too closely.[5] I want to urge that there are really two questions here, not one. In my view, distinguishing between these two questions is extremely important because, while I am persuaded by Clark's answer to the first question, I am not persuaded by his answer to the second.

2.1 The Attractions of Places

Why, then, does Clark want to insist on SP, which identifies sensory individuals with places? As I read him, Clark offers two motivations for SP in [Clark, 2000].

A first reason for endorsing SP is simply that places could (logically speaking) play the role of individuals needed to solve the Binding Problem. And if Clark is right that we need sensory individuals, then taking individuals to be places is one way of answering our need (but see the case of the two mists in §2.2). But, of course, this motivation couldn't seal the case for SP by itself, since it leaves open the possibility that there is some other, better motivated, answer to our need.[6] In §§3–5, I'll attempt to provide just such an alternative.

A second reason for the contemplated identification comes from the spatial localizability of sensory features — perceivers typically believe that the sensory features they perceive are instantiated at locations in space. For example, when you perceive that something is red, you typically hold a belief (except perhaps in unusual circumstances that can be cooked up in the psychophysics lab) that the red something you perceive is in some particular region of the space around you. If we follow Clark in identifying individuals with places, then the same answer we give to the Binding Problem will also explain why perceivers bind features to locations — in effect, we will have reduced two mysteries to one. This motivation for identifying sensory individuals with places seems to me to come out in passages like the following, where Clark is considering a version of the Binding Problem that involves tactile perception of two spatially separated stimuli on the skin:

> Some of the mystery ... is wrapped up in issues of convergence and binding: of how the nervous system manages to represent two points

> as two, and how one might discriminate among pairs of stimuli with differing combinations of features. But part of the mystery remains centered on the spatial character of the experience. How does one manage to feel the touch *there*, at that particular point (46)?[7]

At this point we should distinguish two separate phenomena so that we can put one of them aside. One phenomenon (ultimately not our target) is this: sensory features are located (and believed to be located) in spacetime. Another phenomenon (more relevant for our purposes) is this: particular sensory features are located at (and believed to be located at) particular spacetime regions.

The first phenomenon can be bracketed because, while interesting and significant, it has very little to tell us about sensory individuals *per se*. For notice that the same thing seems to be true of non-sensory features. For example, I take it that *being water* is not a sensory feature: exemplifying *being water* is a matter of having the right hidden essence — something not available to perception. All instantiations of this property occur at spacetime regions as well, and are (I take it) generally believed to occur at spacetime regions. But no one would suggest that the individuals that exemplify *being water* are places. On the contrary, the natural response is that this property — or any other property, for that matter — is exemplified in spacetime regions because properties are exemplified by individuals, and that individuals are necessarily located in spacetime regions. This line of thought suggests that the first question about the localization of sensory features is a special case of the much more general truth that property instances, *qua* individuals, are necessarily located in spacetime regions. This interesting metaphysical generalization surely deserves some sort of explanation (to take some famous examples, both Aristotle and Kant had lines to sell on this problem). But, just because the problem is so general, it would be misleadingly specific to take the spatial localization of sensory features to bear on the nature of sensory individuals in particular. This is why I don't believe that the first phenomenon is relevant to our purposes.

The second phenomenon, which I take to be Clark's main quarry, is that particular sensory features are located at particular spacetime regions. Since I'm going to argue that sensory individuals are not (in general) places, I want to acknowledge immediately that the phenomenon under consideration here is important and deserves an explanation. Moreover, if SP provided the only possible explanation of this phenomenon, I would join Clark in taking it to be a powerful motivation for SP. However, my contention in §4.1 will be that, on the contrary, the phenomenon can be explained without the identification of sensory individuals with places.

To sum up, while Clark gives us some suggestive reasons for construing sensory individuals as places, nothing in what he says rules out the possibility that there may be other, and more plausible, ways of understanding sensory individuals. Before I go on to argue that this possibility is actual (§§3–5), I want to point out what seem to me some serious costs associated with the view of sensory individuals that Clark proposes.

2.2 Troubles With Places

Having argued that the motivations in favor of SP are not decisive, I now want to urge that the view comes with a number of substantial difficulties. Some of these difficulties are, in my view, more serious than others; but, taken together, I hope they will weaken the appeal of SP, and that this will provide motivation for the alternative answer to the Binding Problem I shall offer in §3.

What, then, is wrong with taking sensory individuals to be places?

To begin, it is worth noting that SP is fairly revisionary. When you ask me what, in my immediate environment, exemplifies a sensory feature such as *redness*, I am likely to mention the ripe raspberry on the table, the book on the desk, and the like. I would not, in ordinary circumstances, answer that a certain region of space exemplifies *redness*.[8] Now, this consideration is certainly not decisive by itself: there is no reason to expect that our naive views about the bearers of sensory features get things right, and it certainly would not be the first time that our naive views about perception got things wrong. On the other hand, it is only fair to acknowledge that SP really does amount to a revision of the naive view, and therefore incurs the onus of explaining why the latter seems so plausible.

A second difficulty with SP is that, despite what was claimed in §2.1, I doubt that this position can solve all the instances — even all the visual instances — of the Binding Problem after all (cf. [Matthen, 2004]). Before I come to some actual cases that make this point, I'll creep up to the point by constructing an imaginary case of the sort that I need and showing why it is damaging to SP. The sort of case I have in mind is one in which there are two objects located at at one and the same spatial region, and such that perception binds one set of features to the first object and a second set of features to the second object.[9] For example, suppose we have a red, fine mist in spacetime region R_1 and a green, granular mist in neighboring but distinct spacetime region R_2; so far, there is no trouble for Clark's view since we are imagining that there is only one set of properties per region of spacetime. But now imagine that we move the two mists into a common region R_3 of spacetime, where they are uniformly distributed. Can visual perception solve the Binding Problem within R_3? Can it distinguish between sensing a red fine mist and green granular mist, on the one hand, as against sensing a red granular mist and a green fine mist, on the other? If it can, then perception can bind *redness* and *fineness* to one entity and *greenness* and *granularity* to another entity. But in the imagined case the entities in question can't be spacetime regions, since the spacetime region in which the first two properties are instantiated is identical with the spacetime region in which the second two properties are instantiated (viz., R_3).

Now, a defender of SP might respond to this case by urging that the two mists cannot share the single spacetime location R_3; on this line, the case just shows that the relevant perceptual features are instantiated in very small sub-regions of R_3, such that no two sets of such features can be co-instantiated in R_3. But why should we accept this objection? Notice that there didn't seem anything objectionable about taking *redness* and the other sensory features to

be exemplified in large, whole-mist-sized spatial regions R_1 and R_2 before we combined the two mists in R_3; it is unclear why moving to R_3 or combining the two mists should make for any relevant change in this situation. Consequently, I don't find this concern a substantial obstacle for the case.

A more pertinent worry about the challenge I'm mounting concerns whether perception can indeed solve the Binding Problem in such cases. Unfortunately for SP, recent results indicate that it can. For example, [Blaser et al., 2000] superimposed two distinct Gabor patches at a single spatiotemporal region while smoothly varying the orientation, spatial frequency, and color of each patch (see figure 1).[10] They showed that subjects could distinguish between the two

Figure 1: Parts 1 and 2 depict single Gabor patches; each has a value for the parameters of orientation, spatial frequency and color. Part 3 is a static representation of the stimuli used in [Blaser et al., 2000]: here the two single Gabor patches have been superimposed in a single spacetime region. (Used by permission.)

spatiotemporally coincident patches, and could even track the two separately as they moved through "feature-space" (but stayed fixed in locational space).[11] But this means that subjects can bind together the features of each patch and distinguish them from the features of the other patch: as it might be, patch$_1$ is (at time t) oriented at 20° from vertical, has a spatial frequency of 2 Hz, and a saturation of 45%, while patch$_2$ is (at time t) oriented at 40° from vertical, has a spatial frequency of 6 Hz, and a saturation of 15%. Solving this (visual) instance of the Binding Problem, then, requires two distinct individuals, but occurs within a single spatiotemporal region. Consequently, as in the imagined case of the coincident mists, a commitment to SP allows for too few sensory individuals for the (in this case, actually observed) solution of the Binding Problem.[12]

As in the imaginary case of the two mists, an opponent might deny that the result just discussed shows that there are two sensory individuals at one spatiotemporal region. Instead, she might insist, the result shows only that the relevant sensory features are instantiated at very small subregions of the shared region within the Gabor patches are located. But once again, there seems little reason to accept this objection. After all, when there is only a single patch in a given region, there is no difficulty in taking the relevant sensory

features to be exemplified by the whole patch; and no reason has been offered for thinking that adding a second patch to the region should shrink down the size of sensory individuals. Moreover, note that the expedient under consideration is unavailable for at least one of the sensory features at issue: the spatial frequency of a Gabor patch often can't even be defined for subregions sufficiently small to avoid spatiotemporal overlap (unless one has further information about the waveform, which naive perceivers don't). For these reasons, the objection under consideration seems not to block the problem I have raised for SP: there are (visual) instances of the Binding Problem that are solved by perception, but for which the view cannot account.

This brings me to a third problem for SP concerning the possibility of tracking a single object as it moves through spacetime. On an ordinary description, we would say that an object moving through space retains many of its (perceptually available) features — its color, shape, and size, for example. But, on a story like Clark's, perception must represent that these features are instantiated first by one location, then another, then another, and so on. In a recent review, Susanna Siegel wonders about this point:

> What happens in sensory phenomenology when a subject sees a basketball make its way from the player's hands to the basket? The information that it's one and the same basketball traversing a single path is not given by sentience if sentience is limited to feature-placing. On Clark's view, ... [this information] has to be given non-sensorily. The subject's visual experience stops short ([Siegel, 2002], 137).[13]

This means that, at the very least, defending SP requires saying what extra-perceptual mechanisms facilitate visual tracking (e.g., in subjects as young as 1-month-old; cf., [Kremenitzer et al., 1979], [Maurer and Young, 1983], [Johnson et al., 1991]).[14]

A final (and, to my mind, most compelling) cluster of problems for SP concerns issues about the semantics of natural language predications involving demonstratives. Clark explicitly ties his views about perception to semantic issues in [Clark, 2000], chapter 4, where he argues that the sensory individuals posited to solve the Binding Problem play an essential role in perceptual demonstration. In particular, since he thinks perceptual demonstrations have to single out entities that are available to the perceptual system, he holds that the demonstrata of perceptual demonstrations just are the entities that bear sensory features — viz., on his view, places. And, since the demonstrata of perceptual demonstrations serve as referents for many uses of demonstratives in natural language (so-called perceptual demonstratives), it follows that the referents of perceptual demonstratives will turn out to be places. Thus, when you utter 'this is red' and complete your utterance with a perceptual demonstration (canonically, a pointing), you are attributing *redness* to a certain place that is the demonstratum of your demonstration.

This semantic conclusion seems counterintuitive on its own. However, it is worth noting that the situation is even worse, because these semantic ramifica-

tions spread to NPs other than perceptual demonstratives as well. To see this, consider the following sentences (which we may imagine to have been completed by perceptual demonstrations):

(4) This is a hand and it is my favorite body part.

(5) That is my wife and I love her.

Since the pronomial 'it' in the second conjunct of (4) is anaphorically bound to the subject of the first conjunct of (4), the two share a referent; likewise for the perceptual demonstrative in the first conjunct of (5) and the pronoun 'her' in the second conjunct of (5). Since, on Clark's view of perceptual demonstratives, the referents in question are places, he is committed to holding that I use (4) and (5) to proclaim my affection for places. (Arguments to analogous conclusions can be mounted using other pronouns.) This is, to put it mildly, an unusual way to understand sentences like (4) and (5).

But, even putting this last point aside, there is a more direct reason to be worried about the claim that perceptual demonstratives refer to places — namely, that there are plenty of cases where we use a perceptual demonstration to attribute some property that spatial regions can't exemplify. Consider, as a simpler version of (4), G. E. Moore's famous demonstrative identification, 'this is a hand'. By analogy with the case of 'this is red', it looks like Clark's view would be that the (completed) utterance attributes *handhood* to a place; but this is an obvious category mistake — too obvious, surely, for the likes of Moore. Clark attempts to avoid this absurd consequence by inserting the occupation relation into the analysis. Thus, he reads the (completed) utterance as claiming that a certain place (the demonstratum of the demonstration) is occupied by a hand (142–143).

But this solution, it seems to me, is no less absurd. For one thing, it seems *ad hoc* to claim that the semantics for the predicative VP 'is red' does not involve the occupation relation, while the semantics for the predicative VP 'is a hand' does involve the occupation relation: no motivation has been offered by this move other than the need to save the theory from strange consequences. For another, assuming that what 'is a hand' expresses does not involve the occupation relation when it occurs without a perceptual demonstrative (e.g., if I use a name or a description rather than a demonstrative as my subject NP, as in 'Lefty is a hand'), Clark is committed to the claim that this predicate is systematically ambiguous between what it expresses when combined with a perceptual demonstrative and what it expresses when combined with other sorts of NPs. This concern obviously generalizes to other predicates, so it seems that Clark is committed to saying that all predicates in the language are systematically ambiguous in this way! But the evidence seems to tell against this prediction: for example, we don't find distinct expressions corresponding to each of the alleged readings in other natural languages.

It would appear, then, that there are several seemingly untenable semantic consequences of Clark's claim that perceptual demonstratives refer to places, rather than the objects that occupy places. This claim is counterintuitive on

its face, and it certainly runs counter to all the accounts of the semantics of demonstratives I've ever seen. But notice that the semantic difficulties we're now confronting are, perhaps surprisingly, intimately connected with Clark's main contentions about (not semantics, but) perception. For, if we accept Clark's reasonable suggestion that the referents of perceptual demonstratives must be the entities to which perception (which, after all picks out the referents of perceptual demonstratives) attributes features, then we are forced either to accept the bizarre semantic consequences discussed above or to give up SP.

3 Individuals as Visual Objects

In my view, the problems raised for SP in §2 give us ample reason for withdrawing our support from that claim, and retreating to SI. But this retreat leaves us in need of an alternative understanding of sensory individuals. The alternative understanding that I want to suggest, in this section, is (subject to an important qualification below) roughly continuous with what I take to be an attractively naive view about the bearers of sensory features; viz., that they are (in most cases) the objects — the tables, chairs, basketballs, people, and other objects — that we ordinarily describe ourselves as perceiving.[15]

Now here is the promised qualification: although I think of this proposal as continuous with naive thinking about perception, what I'll really be defending is the systematic empirical elaboration of the idea under the rubric of object-based attention by a number of psychologists and vision scientists.[16] For purposes of expository simplicity, and in order to engage the discussion of [Clark, 2004] directly, I'll work largely with the version of the view developed by Zenon Pylyshyn (mainly in [Pylyshyn, 2001], but see also [Pylyshyn, 1989], [Pylyshyn, 1998], [Pylyshyn, 2000], [Pylyshyn, 2003], and [Pylyshyn and Storm, 1988]) that makes crucial use of a notion of visual objects.

Before I go on to discuss this account and its implications, I want to make one more caveat (on which I'll elaborate in §3.3): I will not be claiming that *all* the bearers of visually extracted features are visual objects, but that (i) many of them are, that (ii) in particular, the best candidates for bearing color and shape properties are visual objects rather than places, and that (iii) appealing to the exemplification of sensory features by visual objects is the right way to resolve typical instances of the Binding Problem that, we've seen, Clark uses to motivate SI.

Our first item of business, then, will be to explicate Pylyshyn's notion of visual objects; this will be the task of §§3.1–3.2.

3.1 Visual Indices

Pylyshyn argues that the visual system provides a "special kind of direct connection between elements of a visual representation and certain token elements in the visual field, a connection that is unmediated by an encoding of properties of the elements in question" ([Pylyshyn, 2001], 128). His proposal is that this di-

rect connection is subserved by a small number (four or five) of "visual indices" that are causally linked to — and thereby allow reference to — distal elements of the visual scene, in something like the way that demonstratives of natural language are causally linked to — and thereby allow reference to — elements of the distal world. Pylyshyn insists that visual indices are "preconceptual," by which he means that they can be bound to an object x not in virtue of the subject's judgment that x falls under certain concepts (say, that x is a table or a person, or that x is red and triangular). On the contrary, this view has it, a visual index is bound to x because x causally activates the subject's visual index mechanisms, prior to and independently of the subject's representing anything about x or its properties. To put the point in other words, if visual indices provide a mechanism for reference to distal items, this is not a kind of reference by description, but reference constituted by some sort of causal connection.[17]

While Pylyshyn marshals an impressively large number of conceptual and empirical motivations for the existence of such visual objects, one of the most important sorts of evidence, and that most prominent in [Clark, 2004], concerns what Pylyshyn calls "multiple object tracking," or MOT:

> In a typical experiment, observers are shown anywhere from eight to 24 simple identical objects (points, plus signs, circles, figure-eight shapes). A subset of these objects is briefly rendered distinct (usually by flashing them on and off a few times). Then all the identical objects move about in the display in unpredictable ways. The subject's task is to keep track of this subset of objects (called 'targets'). At some later time in the experiment (say 10 s into the tracking trial) one of the objects is probed by flashing it on and off. The observer must then indicate whether the probed object was one of the targets.... A large number of experiments ...have shown that observers can indeed track up to five independently moving targets within a field of ten identical items ([Pylyshyn, 2001], 142).

Pylyshyn proposes to explain the phenomenon of multiple object tracking by supposing that the initial flash binds visual indices to the targets, and that the indices stick to their referents as the objects (targets and distractors) move about the display so that the subject can finally indicate whether the probed item was one that was being tracked or not.[18]

3.2 Sensory Individuals as Visual Objects

The brief discussion of visual indices above is enough to suggest an alternative understanding of the (visual) sensory individuals to which SI commits us as what Pylyshyn calls 'visual objects' (and sometimes 'proto-objects') — as whatever can be selected by a visual index. Moreover, Pylyshyn gives an operational criterion for the latter: something counts as capable of selection by a visual index just in case it can be tracked in MOT experiments.

So what sorts of things, on this construal, will turn out to be visual objects? To answer that question fully we need to know what constraints there are on

visual indexing, and this is an ongoing matter of empirical controversy (see, e.g., §6 of [Scholl, 2001] and [Scholl et al., 2001]). But, even prior to the resolution of this controversy, we know some of the things that can get picked out by a visual index — namely, the points, plus signs, circles, and figure-eight shapes used in MOT experiments. Beyond that, on the plausible empirical assumption that subjects can visually track cars, people, basketballs, tables, chairs, and the like in their environments, these objects, too, can count as sensory individuals.[19]

The proposal that I am making, then, is that the sensory individuals cited in SI include the ordinary objects — the tables, chairs, basketballs, people, and so on — that we ordinarily describe ourselves as perceiving. Recall that, on Clark's preferred view SP, sensory individuals are (always) places. In its stead, I am suggesting, we should endorse SO — the thesis that sensory individuals are typically (visual) objects.

3.3 Places As Feature-Bearers in Early Stages

If, as I have suggested, the theory of visual indices gives us an alternative to SP, why do I want to allow that spacetime regions may nonetheless be the bearers of at least some sensory features? The best reasons for this concession concern the exemplification of sensory qualities at early stages of visual processing that occur prior to the assignment of indices.

According to Pylyshyn, the process of assigning visual indices requires, *inter alia*, segregating items from the background ([Pylyshyn, 2001], 145). But, as Clark points out, segregating even a single item x_i from the background is accomplished partly in terms of the extraction of visual features:

> Such segregation requires picking out at least some of the boundaries of x_i and perceiving them *as* boundaries of something distinct from the background.... Even something as simple as edge detection requires registering a "discontinuity of intensity" across spatial intervals; it too is a relation between values at places. And we have not one, but n such figures to segregate from the ground. How could we manage the entire task if *all* the direct reference of visual representation had to be achieved using just four or five visual indices? ([Clark, 2004], 14)

The worry, then, is that the assignment of visual indices requires prior segregation, that (in at least some cases) segregation involves edge detection, and that edge detection requires the assignment of feature values (viz., luminance intensity values) to more entities than we have visual indices.

An initial response to this worry would point out that, even if the luminance features at issue are not exemplified by visual objects, it's not obvious that there's any Binding Problem at this stage, so it might turn out that the visual system gets by without assigning these features to any sensory individuals at all. On the other hand, as Clark points out, luminance discontinuities must be represented as a relation between multiple spatial intervals. It won't suffice to register one luminance value after another as a simple luminance-extractor

would (cf. §1); rather, the visual system needs to represent that the different luminance values hold of different entities that stand in the right relation to count as an edge. Once again, it looks as if the visual system must represent that these features are exemplified by individuals; but in this case, there's reason to think that the individuals can't be visual objects — both because at this stage individuals aren't available, and also because there aren't sufficiently many of them. As usual, it would seem unwise to take the individuals as proximal entities (e.g., receptive fields, retinal clusters): it is some bit of the world, rather than the visual system itself, that bears a luminance value. In the absence of any other reasonable suggestions, it seems plausible (and compatible with the usual descriptions of such early processing by vision scientists) to think that the individuals bearing such features are spacetime regions.

I am prepared to admit, then, that the individuals represented by visual perception as the bearers of at least some visual features are spacetime regions some of the time. Nonetheless, I want to suggest, this concession won't give the fan of SP enough of what she wants; consequently, despite the caveat, I think that SO remains an interesting and substantive alternative to SP.

For one thing, the concession that spacetimes may bear features in some limited cases is certainly insufficient to warrant SP if the latter is offered (as it was in [Clark, 2000]) as a universal claim about the bearers of (visual) sensory features (in fairness, the position of [Clark, 2004] is considerably weaker). On the contrary, the problems urged against SI in §2.2 continue to provide strong reasons for thinking that, in many cases, sensory individuals cannot be understood as spacetime regions: in particular, this view seems to be unacceptable in cases where we perceptually track items through spacetime, in the case of the two mists and the co-located Gabor patches of [Blaser et al., 2000], and in cases where the sensory individuals are the referents of perceptual demonstratives. Moreover, the motivations for the concession to SP I am making apply only to the bearers of visual features that are extracted early in the process of assigning indices; as such, the scope of the concession is (so far) quite small.

Since I'm prepared to concede ground to SP over the cases involving sensory individuals at this early stage, I will be construing the disagreement in what follows as a disagreement about the sensory individuals at later stages, once visual objects are available as an alternative to spacetime regions. In particular, this will mean that the dispute applies to the sensory individuals that bear such paradigmatic visual features as shapes and colors that [Clark, 2000] appealed in the course of arguing for SI, and thereby raised the question about the nature of sensory individuals in the first place. I'll argue that SO provides a more satisfactory understanding of visual perception in these cases.

4 Objects and Places: An Economic Assessment

So far I've argued that, while we should join Clark in accepting SI, we should reject his further thesis SP in favor of SO — we should understand the sensory individuals to which SI commits us as visual objects, rather than as places (in all

but a limited number of cases that occur very early in visual processing). Since, at this point, we have before us two different proposals for understanding sensory individuals, we are now in a position to compare the two. I'll engage in that comparison in this section by trying to tally up the costs and benefits of each account. My contention will be that SO retains the benefits of SP (§4.1) while avoiding its costs (§4.2). I'll complete this assessment in (§5) by considering some further possible costs of SO. In the final accounting, I'll suggest, SO is the better buy.

4.1 Visual Objects and the Benefits of Places

As we saw in §2.1, Clark has two main motivations for SP. The first is that places could (logically speaking) play the role of individuals needed to solve the Binding Problem. And second, SP comes with a natural explanation of the interesting fact that sensory features are instantiated at (and subjects believe that they are instantiated at) at locations in space.

As I've already noted, the first motivation is only persuasive in the absence of alternative explanations of the Binding Problem. But now that SO is on the table, an alternative explanation is ready to hand. To return to the instance of the Binding Problem considered in §1, a subject whose perceptual capacities include the capacity to represent visual objects as the bearers of sensory features (by having visual objects selected by her visual indices) can thereby represent that both *redness* and *triangularity* hold of the very same entity:

(1″) visual object o_1 is red,

(2″) visual object o_1 is triangular.

In this way, such a subject could recognize that first feature is bound to the second with respect o_1 (and not with respect to other visual objects), and could thereby distinguish the red triangle from the red non-triangles and non-red triangles in the visual scene. Adopting SO rather than SP, then, allows a solution to the Binding Problem that initially inspired SI.

What of the localization of sensory features? Recall that what matters for our purposes is that particular sensory features are bound to particular spacetime regions. How can the proponent of SO explain this important phenomenon? According to SO, bearers of sensory features are whatever can be selected by a visual index. But, as we noticed, it is a general metaphysical truth, not denied by anyone in the present debate, that individuals (*qua* bearers of properties, including but not limited to sensory features) are necessarily located in spacetime regions. In particular, all the things that can be selected by a visual index must be located at spacetime regions, since what it is to be selected by a visual index just is to be tracked visually through regions of spacetime in MOT experiments. When an visual object exemplifies some particular sensory feature, then, it exemplifies that feature at a particular spacetime region (viz., that spacetime region it happens to occupy while exemplifying the feature). It is this fact, according to the proponent of SO, that explains the spatial localization of sensory features.

Of course, the latter explanation of the phenomenon differs in character from that offered by the champion of SP: it works not by supposing that the bearers of sensory features are spacetime regions, but by holding that the bearers of sensory features, like all other individuals in our world, occupy spacetime regions. But I don't see why this should count against the explanation.[20]

4.2 Visual Objects and the Costs of Places

I've argued that SO captures the benefits used to motivate SP. In addition, SO avoids the costs adduced for SP in §2.2 — that the view is (i) excessively revisionary, (ii) insufficient to account for perceptual tracking of objects through spacetime, (iii) unable to explain all instances of the Binding Problem, and (iv) yoked to a cluster of unattractive semantic consequences. I'll consider these in turn.

First, the proponent of SO takes the sensory feature *redness* to apply to visual objects, which can include ripe raspberries and books rather than places. As such, SO is unlike SP in that it agrees with the naive view (at least, in very many cases) about the sorts of things that exemplify sensory features. For this reason, SO is far less revisionary than SP.

Second, it should be clear that SO is especially well-suited to the explanation of perceptual tracking of objects through spacetime. Indeed, according to the proponent of SO, sensory individuals just are those things that can be tracked through spacetime. Siegel's problem of the basketball — a serious obstacle for SP — is no problem for SO.

Third, SO is up to explaining the potentially troublesome instance of the Binding Problem involving two co-located mists. The difficulty here was one of perceptually binding *redness* and *fineness* to one mist, while binding *greenness* and *granularity* to a second mist, after the two mists had come to occupy one and the same location R_3. As we saw, the proponent of SP lacks the resources to explain this case of perceptual binding (assuming it is a genuine case of perceptual binding!), insofar as her view makes available only a single sensory individual — viz., R_3. On the other hand, suppose we follow SO in taking sensory individuals to be visual objects. Given our operationalization of visual objects as whatever can be selected by a visual index, we can take each mist as a separate visual object. Then we will say that (i) *redness* and *fineness* are bound to one visual object that starts out in region R_1 and then moves into region R_3, while (ii) *greenness* and *granularity* are bound to a second visual object that starts out in region R_2 and then moves into a region R_3. Significantly, this view allows for a distinction between the two visual objects even during periods while the objects occupy a single region of spacetime (see note 18). Hence, the spatiotemporal coincidence of the two mists is no obstacle to construing them as distinct visual objects; and this, in turn, allows us to explain the property bindings in the case of the two mists as just another ordinary instance of the Binding Problem. (*Mutatis mutandis* for the perception of spatiotemporally overlapping Gabor patches from [Blaser et al., 2000].)

Finally, we saw in §2.2 that SP carried with it a number of unlovely consequences concerning the semantics of demonstratives, and that these resulted in further objectionable views about the semantics of other expressions (that all predicative VPs in the language are systematically ambiguous, that pronouns anaphorically bound to perceptual demonstratives refer to places, etc.). For, on the (reasonable) assumption that the referents of perceptual demonstratives are the entities that bear sensory features — viz., sensory individuals, a commitment to SP amounts to saying that the referents of perceptual demonstratives are places. On the other hand, if we follow SO in holding that sensory individuals are visual objects, then we end up instead with the conclusion that the referents of perceptual demonstratives are (as one might have hoped) tables, chairs, and hands, *inter alia*. This conclusion both avoids the further semantic problems adduced for Clark's view, and agrees with the usual semantic treatments of perceptual demonstratives. I take these to be strong grounds for preferring SO over SP.

5 Further Costs of Objects?

Even if I am right that SO enjoys the benefits of SP while avoiding the costs examined so far, it is worth wondering whether the view carries further costs — costs not accruing to SP — that make SO less attractive than its rival.

5.1 Location-Based Attentional Selection

A first group of objections to SO comes from evidence for location-based cuing of attention. Clark reminds us of several findings that have been used to argue that perceptual attention can be bound to regions of spacetime rather than objects — findings concerning spatial cuing of attention, auditory localization, and cross-modal cuing.

I want to take issue with Clark's interpretation of these findings; I'll urge that they are compatible with either Clark's preferred interpretation that attention is bound to locations or the contrary interpretation that attention is bound to objects at those locations. Before I come to this, however, I want to point out that, even spotting Clark his interpretation of the spatial cuing results, this still wouldn't give Clark what he needs to support SP. SP demands that visual perception represents the exemplification of sensory features by locations. But even granting that attention is bound to locations, it doesn't follow that those locations bear sensory features, much less that visual perception represents the exemplification of sensory features by locations. Indeed, theorists typically motivate the view that attention binds places by finding effects associated with attentional binding (e.g., enhanced speed and accuracy in the recognition of subsequently presented items in nearby spatial regions) in cases where, as they would put it, there is no obvious sensory feature exemplified by anything. The whole point of these cases is to find attentional effects in empty space, where 'empty' has to entail the absence of sensory features, lest an opponent claim

that there is an object for attention to bind after all. But if there are no sensory feature exemplifications, then *a fortiori* there are no sensory features exemplified by locations. The upshot, then, is that even if there were decisive evidence that attention binds locations, this would give no support, by itself, to SP.

Having said that, I now want to argue that the spatial-selection experiments Clark mentions needn't be taken to show that attention binds locations, but rather can be understood as showing that attention binds objects that are at locations.[21]

The first sort of evidence Clark points to in this regard comes from spatial cuing results ([Posner, 1978], [Posner et al., 1980], [Posner et al., 1987], [Posner and Petersen, 1990], [Posner and Rothbart, 1992]) showing that subjects are faster and more accurate at identifying targets at locations cued by an arrow pointing to the side of the screen where the target would appear but prior to the appearance of the target, and slower and less accurate if the cue points to the side opposite that where the target appears. As Clark sees things, "the problem here posed for object-based accounts is to explain how the facilitation or inhibition endures in the interval after the cue stops and before the target appears. There is no obvious object one can nominate to serve this role" ([Clark, 2004], 15).

But the proponent of object selection has a ready response to this challenge. Namely, she can claim, the object that binds attention in this interval is the edge of the monitor on which the stimuli are presented. The proposal would be that the cue first directs attention to either the left or the right monitor edge (depending on the direction in which the cue arrow points), and that accuracy and speed are enhanced when the target appears in a position closer to the attended monitor edge (and inhibited when the target appears in a position further from the attended monitor edge) since it takes longer to shift attention between visual objects at disparate spatial locations.[22] (Similar explanations can also provide object-based accounts of effects such as inhibition of return and the facilitation of intermediate locations that Clark adduces in support of spatial selection ([Clark, 2004], 16).)

Next consider a case of auditory localization: the subject hears a telephone ring, and localizes the sound at a region of space. Clark thinks this case tells against object selection views because it may turn out that there is no object bearing the relevant auditory features that can serve as a locus of auditory attention; this will happen when the subject has been tricked into mis-locating the sound by an auditory hallucination.[23] It will be helpful to consider a veridical (non-hallucinatory, non-illusory) case of auditory localization first. Suppose, then, that there *is* a sound source (say, a telephone) at the region localized by the subject; it is natural to say here that the subject binds attention to the sound source. In short, the case is aptly described in terms of object selection by auditory attention. Now, however, suppose there is no sound source at the indicated spatial region, but that the subject undergoes an auditory hallucination. Is this, as Clark claims, "an 'object-free' indication" ([Clark, 2004], 16)?

It's hard to see why. As Clark points out in another context ([Clark, 2004], 22), anyone who allows for the possibility of hallucination is committed to ac-

cepting merely intentional bearers of sensory features (given the very definition of 'hallucination'; see note 23). Therefore, on the further (concessive) assumption that the bearers of sensory features are identical to the loci of attentional selection, proponents of both object selection and spatial-selection who countenance hallucination must posit some sort of merely intentional loci of attention: presumably the fan of object selection will posit merely intentional objects, while the fan spatial-selection will posit merely intentional places. But if this much is allowed, then auditory hallucination poses no trouble for object selection after all; either one will refuse to recognize merely intentional auditory objects, in which case there are no cases of auditory hallucination to raise trouble in the first place, or else one will accept merely intentional auditory objects, in which case auditory hallucination can be straightforwardly understood as an instance of object selection by a merely intentional auditory object.[24]

In fact, although Clark presents auditory hallucinations as special problems for the proponent of object selection, hallucinations are arguably much more troublesome for a spatial-selection theorist. As we've seen, realism about hallucination together with the premise that the bearers of sensory features are attentional loci requires merely intentional loci of attention. However, while philosophers have famously puzzled about how to understand the merely intentional, it is even more puzzling how there could be merely intentional places: there don't seem to be empty holes in space into which a merely intentional place could go, so it's hard to see how spatial-selection could ever select a merely intentional place rather than a *bona fide* (not merely intentional) place. The worry, then, is that a proponent of spatial-selection would be forced to redescribe all ostensible hallucinations as illusions. The upshot is that, far from defeating object selection, hallucination cases provide *prima facie* reason for thinking that spatial-selection can't be the whole story about perceptual attention.

Turn now to a third sort of evidence Clark adduces in favor of spatial-selection — namely, those in which localized auditory cues serve to cue visual attention ([Clark, 2004], 17–18). Once again, I think the data here are more friendly to objects than he allows. A proponent of SO can describe these cases as follows. When the auditory cue occurs, the subject attributes auditory features to some auditory object and localizes it to some spacetime region R_A. Next, the auditory object ceases to bear auditory features; still, there's no reason for thinking that the object itself fades away during this interval. Finally, a visual target occurs, and the subject both attributes visual features to this visual object, and represents it as occurring at some particular spacetime region R_V. Why will the subject attending to an object at R_A will attach an index to a new perceptual object (of whatever modality) more quickly and accurately if that new object is closer to R_A? Not because the region R_A becomes the locus of attention in the interval between the occurrence of the auditory and visual objects, but because the auditory object persists at R_A despite ceasing to bear features, and attention is moved between objects as a function of their propinquity.

In sum, much of the evidence for location-based attentional selection can be explained in a way that is compatible with object-based selection. More-

over, I've argued, even if all attentional selection were spatial-selection, this still wouldn't give us reason for thinking that perception represents visual properties as holding of spacetime points/regions rather than visual objects; consequently, even on this concession the results at issue fail to provide support for SP over SO.

5.2 Visual Objects and Objects Proper

A final concern about SO is that what it counts as visual objects don't satisfy standard ontological criteria for objecthood. Recall that, for the proponent of SO, a visual object is understood as anything that can be selected by a visual index, where this last criterion is operationalized in terms of subjects' capacity to track elements of visual displays in MOT experiments. But this seems to amount to something different from our ordinary notion of objects. After all, at one point Pylyshyn himself claims that visual objects are "precursors to" and "less than" tables, chairs, and other "real physical objects": "The concept of a 'proto-object' is a general one that has been used by a number of writers ...in reference to clusters of proximal features that serve as precursors in the detection of real physical objects. What these uses have in common is that the refer to something more than a localized property or 'feature' and less than a recognized 3D distal object" ([Pylyshyn, 2001], 144, note 5; cf. 154). Moreover, while there remain open empirical issues about how far this criterion for visual objecthood extends, we know that the criterion will be satisfied by many things that we might be reluctant to count as objects — that it will apply to things that are not "objects proper" ([Clark, 2004], 21ff), including the merely intentional (22), and perhaps things like Marr-Hildreth edges and ocean waves (since these can be visually tracked). But, Clark objects, this just shows how far visual objects are from being objects proper.[25] For example,

> ...waves are still quite a ways from being "objective particulars". Consider: Two waves can fuse; one wave can split into two; "a" wave has entirely different parts at different times; and if all those parts are arranged in just the same way at a different time, it is a purely verbal question whether we have the same wave again, or a different one that is qualitatively identical. Waves are often mentioned by Buddhist thinkers who want to point to something that appears to be an object but is not. In this I think they are right ([Clark, 2004], 24).

Now, it is true that the visual objects countenanced by SO might fail to count as objects on standard ontological criteria for objecthood. But I fail to see that this is a problem for the view. The notion of visual objects to which SO appeals was proposed to explain certain perceptual phenomena, such as (principally for our purposes) the perceptual capacities for solving the Binding Problem and for tracking in MOT experiments. Given that the puzzles motivating the introduction of visual objects are manifestly perceptual rather than ontological, there is no reason to expect visual objects to deliver the ontological goods, and

there is no reason to spurn visual objects if they do not. To test the contention that ocean waves are visual objects, then, the pertinent questions are perceptual (e.g., can we solve a Binding Problem involving distinguishing a tall and dark wave from various tall non-dark waves and non-tall dark waves?) rather than ontological (e.g., can there be fusion and fission involving waves?). Beyond this, in order to preserve some link to the naive understanding of objects as the bearers of sensory features that inspired SO, we should hope that visual objects include at least some of what we would naively count as objects — to take some examples that have come up in the foregoing, we might hope that basketballs, people, chairs, and tables would fall under both headings. Luckily, as far as we can tell so far (and despite Pylyshyn's caution in the passage quoted above), it seems that they will.[26]

6 Conclusion

As I read things, there are two core claims at the heart of [Clark, 2000]: first, that we cannot understand perception without taking it to represent individuals as the bearers of sensory features (SI); and second, that the individuals in question are best conceived of as regions of spacetime surrounding the perceiving organism (SP).

The first of these claims, SI, seems to me an extremely important insight that has numerous and deep implications for our understanding of perception. However, I am far less confident about Clark's second claim, SP. Since Clark believes in both SI and SP, his discussion of these positions and the arguments for them in [Clark, 2000] sometimes fail to separate cleanly between the two. A first goal in this essay, then, has been to urge that the two theses should be disentangled and evaluated separately.

A second goal has been to carry out the needed evaluations. While I have agreed with Clark's arguments in the service of SI (§1), I have suggested that his arguments for SP are less convincing (§2). Moreover, I have suggested, there are good reasons for rejecting the latter in favor of SO — the view that the individual bearers of sensory features are best conceived as (visual) objects (§§3–5). To be sure, there remain many unresolved issues about the notion of objects appealed to here that deserve (and are receiving!) attention from those studying perception. Nonetheless, it seems to me that (visual) objects provide a plausible and attractive alternative way of understanding the sensory individuals whose importance Clark has so eloquently demonstrated.[27]

Notes

[1]Clark notes that the 'Many Properties' problem of ([Jackson, 1977], 65) is an early philosophical version of this puzzle, and that it also goes under the name 'the feature integration problem' in the psychological literature.

[2]Objection: Suppose Clark is right that the Binding Problem cannot be solved by a system that lacks the ability to pick out individuals. Why not suppose that perceptual systems pick out individuals through conjunctions of features rather than by anything like singular terms? If this is possible, then the binding puzzle that Clark has brought to our attention could be solved by a set of feature extractors after all (albeit perhaps only a large and complicated set of feature extractors). For example, we could exchange ($1'$), ($2'$), and (3) for more complicated quantificational premises that involve only the representation of features:

(1″) Red $(\iota x(F_1x \wedge F_2x \wedge \cdots \wedge F_nx))$.

(2″) Triangular $(\iota x(G_1x \wedge G_2x \wedge \cdots \wedge G_mx))$.

(3″) $\iota x(F_1x \wedge F_2x \wedge \cdots \wedge F_nx) = \iota x(G_1x \wedge G_2x \wedge \cdots \wedge G_mx)$.

Response: As Pylyshyn has pointed out, the quantificational scheme represented by ($1''$), ($2''$), and ($3''$) can get to be very unwieldy, and is subject to serious psychological concerns about chronometry as well as the integration of perception and action ([Pylyshyn, 2003], chapter 5; cf. [Pylyshyn, 2001], 137–138). More significantly for present purposes, the scheme we are now considering actually concedes Clark's main point, since it does *not* involve only the representation of features. On the contrary, the bound variable here serves to represent an individual, and it is crucial that the features F_i in ($1''$) and ($3''$) (and the features G_i in ($2''$) and ($3''$)) are all predicated of one and the same individual. Hence, the quantificational scheme under consideration is not a way of avoiding a commitment to sensory individuals, but an alternative (quantificational) way of representing them.

[3]As Clark himself points out, it is not clear that the argument under consideration, developed mostly with respect to vision and touch, generalizes to other modalities. Clark puts this matter this way:

> Perhaps human olfaction . . . [lacks subject-predicate structure]. Can one smell two distinct simultaneous instances of the same acrid odour? Can one distinguish a presentation in which something smells both acrid and musty from one in which something smells acrid and something *else* smells musty? If the modality lacks such [subject-predicate structure], then one cannot solve that modality's variant of the Many Properties problem. In such a case the flux of

> odours present at a given time could indeed be adequately described with propositional logic alone (79).

It seems to me that the only sensible course of action is to investigate the modalities empirically (however many there are), one at a time, to see whether the perceptual system really solves the Binding Problem for each of them. Only for those modalities in which the problem is solved will Clark's argument justify us in claiming that those modalities have subject-predicate structure. It may be, then, that what Clark is ultimately offering us is better regarded as a theory of visual and tactile sentience, rather than a theory of sentience *tout court*.

[4]

There are temporal versions of the same Binding Problem to which SP is offered as a solution; see, for example, [Yamamoto and Kitazawa, 2001], [Shore et al., 2002]. Assuming Clark would want to treat these analogously, SP should really amount to the identification of sensory individuals with points/regions of spacetime, rather than simply points/regions in space. Having noted this point, I won't be picky about it in what follows.

[5][Matthen, 2004] emphasizes this point as well.

[6]Although this is nowhere explicit in [Clark, 2000], I believe that Clark failed to take this possibility seriously because, when he was writing the book, he couldn't imagine anything else that is (i) psychologically plausible, and (ii) up to the job of solving the Binding Problem. On the other hand, since writing the book, Clark has obviously thought extensively about the alternative proposal I'll be defending, and has even written about the topic in [Clark, 2004].

[7]Notice how Clark moves between the Binding Problem question in his first quoted sentence, and the spatial location question in the second and third, by asserting that they are distinct parts of a single "mystery." It is just this sort of transition that I want to be on guard against in what follows.

[8]Of course, sense-data theorists would more readily attribute *redness* to spatial regions. But this feature of the view was a quite deliberate theoretically-informed attempt to overturn the naive view that attributes *redness* to objects; as such, the presence of a similar commitment in sense-data theory in no way diminishes the counterintuitive flavor of SP. (An important distinction between Clark and sense-data theory on this issue is that Clark's places are locations in the three-dimensional physical space around the perceiver, while sense-data theorists notoriously took the relevant places to be elements in a distinct, mental space; see the discussion in [Clark, 2000], chapter 3 on this point.)

[9]The idea of perceptual binding involving multiple co-located stimuli goes back at least to Ulric Neisser's work on 'selective looking' in the 1960s and 1970s; cf. [Neisser, 1967], [Neisser, 1979], [Neisser and Blecklen, 1975].

[10]A Gabor patch — a commonly used stimulus pattern in vision science — is a local patch of a Gabor function; a Gabor function is obtained by multiplying a sinusoidal luminance grating by a Gaussian weighting function. The spatial frequency of a Gabor patch is identical to the spatial frequency (measured in cycles per degree of visual angle) of the underlying sinusoidal grating.

[11]

Similar results are reported in [Valdes-Sosa et al., 1998b] and [Valdes-Sosa et al., 1998a].

[12]Objection: Perhaps subjects perceive the two (two-dimensional) patches at different depths, and so locate them in distinct spatiotemporal regions after all.

Response: [Valdes-Sosa et al., 1998b] give two reasons for doubting this explanation. First, subjects in their similar version of the binding task in a single two-dimensional location "reported that they perceived the displays as quite flat" (B20). Second, in a companion study, these investigators used binocular disparity to create a much more salient difference in perceived depth between the two patches; this manipulation did not improve performance in distinguishing and tracking the patches.

[13]

Strawson raises similar concerns for a view very similar to SP ([Strawson, 1959], 222ff). He concludes that a scheme like SO is to be preferred on the grounds that its description of cases where particulars move through spacetime (e.g., the case of the moving basketball) will be far simpler than that afforded by SP: "Given our actual situation, and given that we wish to say things having approximately the force of the things we actually do say, then the premium on the introduction of ordinary concrete particulars [as opposed to the introduction of placetimes as concrete particulars] is enormous, the gains in simplicity overwhelming. But this is hardly surprising" ([Strawson, 1959], 225).

[14]I'm not sure what the extra-perceptual mechanism would be, but at least one *prima facie* alternative involving scanning of focal attention is ruled out in [Pylyshyn and Storm, 1988]. These authors showed that tracking cannot (in general) be explained by a serial strategy wherein the subject scans to find the location at a first moment, encodes the location, then scans nearby the stored location to find the updated location at the next moment, and so on. Making liberal assumptions about scanning rates, they showed that the strategy described could not achieve the success rates achieved by subjects in multiple object tracking experiments (about which, see more below).

[15] Here and in what follows I'll restrict myself to discussing the sensory individuals for vision, ignoring other modalities except where noted. One reason for this limitation is that, as discussed above, I am less confident of the arguments for the existence of sensory individuals in non-visual modalities. A second rea-

son is that the empirical research to which I'll be appealing is far more developed with respect to vision than other modalities.

[16]A number of important contributions to this literature appear in volume 80 (2001) of *Cognition* — a special issue dedicated to the topic. Especially useful is Brian Scholl's overview piece, [Scholl, 2001].

[17]This formulation of the point is intended to evoke analogous views about the reference of proper names (e.g., in [Kripke, 1980]) and natural language demonstratives (e.g., in [Kaplan, 1989] and [Perry, 1979]).) And in fact, some of the empirical support Pylyshyn adduces for the claim about visual indices mirrors the famous arguments for the analogous claims about proper names and demonstratives; in particular, Pylyshyn points to the result that visual indices maintain their referential relations across changes in the properties (including locations) of their referents.

[18] One interesting and significant phenomenon revealed by MOT experiments is that (subject to some interesting constraints) subjects can continue to track objects even when they disappear behind occluders and even when one tracked object occludes another because their paths cross. The latter phenomenon is especially interesting for our purposes, since it suggests that subjects are tracking two distinct sensory individuals even while they coincide in spatiotemporal location.

[19]Objection: Sensory individuals are introduced as the bearers of sensory features, including, paradigmatically, single colors. But ordinary objects that subjects can track — cars, people, basketballs, tables, chairs, etc. — typically exemplify different colors in different regions rather than a single, uniform color. Thus, if we follow the present suggestion in taking sensory individuals to include these ordinary objects, then it looks as if they can't serve as the bearers of single colors.

Response: First, recall that sensory individuals were introduced as the bearers of sensory features in general, and that single colors were offered only as examples of the kind of sensory features that are sometimes extracted by the visual system. Since the ordinary objects under discussion do exemplify other sensory features extracted by the visual system (e.g., shape, size, depth), it seems fair to insist that they can play the role carved out for sensory individuals. Second, our operational criterion for objecthood suggests that in cases in which ordinary objects exemplify different colors in different regions, the visual system has alternatives about what to count as a sensory individual. Suppose a subject perceives a basketball made up of white and orange regions. She can either track single-colored basketball-regions (taking color as a way of discriminating between visual objects), or she can treat the distinct regions as a perceptual group (ignoring differences in color between the regions) and track the whole basketball. What serves as a sensory individual, then, will depend on which features are counted and which ignored by the visual system for the

purposes of object tracking on a given occasion. (Again, the question of what alternatives for tracking are available to visual systems in various conditions is an ongoing matter of empirical controversy.) Notice, however, that on any of these alternatives, what is tracked by the visual system is not a spacetime region (as per SO), but a distal object.

[20]Indeed, since SO gives us a way of construing sensory individuals that is independent of their spacetime locations, it leaves us free to treat spacetime location as just one more feature of (visual) objects, roughly on a par with other features such as colors and shapes of (visual) objects. Thus, on this view, a subject who places the feature *greenness* to x at a particular location can be thought of as binding two features (*greenness* and that location) to x in a routine solution to the Binding Problem.

Objection: If *being located at a spacetime region* is a property of x then it is no longer true that all properties are instantiated at spacetime regions; in particular, it seems that x's property *being located at spacetime region* R is not instantiated at a spacetime location.

Response: x's property *being located at spacetime region* R is indeed instantiated at a spacetime region: namely R.

Counter-response: In what spacetime region is x's property *instantiating at* R *the property of being located at region* R?

Counter-counter-response: In region R. And I'm prepared to carry on with this for a very long time. So long as you are prepared to count such further nestings as further properties that x instantiates, so that you will ask about the spacetime region in which x instantiates them, I will continue to have answers for your questions. I warn in advance that the answers are going to be identical each time, so the game may lose its interest quickly.

[21]I'm not committed to the claim that attention never binds locations. On the contrary, the data so far suggest that attention can bind either locations or objects, depending on extremely fine-grained details of the experimental setup. My point is only that (i) there are good empirical reasons for doubting that spatial-selection can be the whole story about visual attention (cf. [Scholl, 2001], §2 for an overview of these reasons), and (ii) the results Clark mentions leave room for an alternative, object-based interpretation.

[22]It should be possible (in principle) to test this proposed explanation by seeing whether the effect remains in conditions where there are no discriminable visual objects to which visual attention can be bound. In practice, however, it can be very difficult to eliminate all possible visual objects from the display: not only monitor sides, but other objects in the vicinity, and even the facing surface of the monitor must be made indiscriminable from empty regions of space.

[23] As I use the term, a hallucination occurs when there is no x that the

subject perceives, but the subject erroneously takes there to be an x that she perceives, and takes it to bear certain sensory features. Hallucinations are to be contrasted with illusions; an illusion occurs when there is some x that the subject perceives, but the subject errs in the sensory features she attributes to x (i.e., (i) the subject perceives x to exemplify properties that, in fact, x fails to exemplify, or (ii) the subject perceives x to fail to exemplify properties that, in fact, x exemplifies).

[24]Clark also brings up pain perception as an example of object-free indication. I'm prepared to concede this (controversial) point for the sake of argument; as pointed out in note 15, I'm not committed to the view that anything like object selection or SO extends plausibly to other perceptual modalities. Indeed, the present digression about audition violates my stated policy of talking only about vision; my justification is that Clark uses the case to draw conclusions about visual attention and visual perception, and so it demands an answer.

[25]Clark also worries that the system of reference to visual objects by visual indices is unlike the system of reference to objects in natural language in several respects (23). But it's hard to see why this should matter. It is surely to be expected that perception and natural language differ in the referential devices they make available; but Frege taught us long ago that differences in the referring apparatus is compatible with identity at the level of what is referred to by the apparatus.

[26]I'm not sure whether this marks a significant departure from Pylyshyn's views, or whether he was just being more cautious than I think he needed to be in the quoted passage. For one thing, it would be very surprising if the mechanisms at work in MOT experiments were wholly unconnected with those at work in tracking distal 3D objects; surely the interest in the former mechanisms derives largely from the hope that they are identical with the latter mechanisms. After all, the point of these experiments is to discern the character of our cognitive/perceptual capacities in general, not just the character of the cognitive/perceptual mechanisms at work in the lab. Finally, I should report that, in personal communication, Pylyshyn has thrown the caution expressed in the footnote to the wind, and has been (tentatively) willing to endorse the view that ordinary 3D distal objects can be visual objects.

[27]I am indebted to Craig Callender, Austen Clark, Alex Holcombe, Mohan Matthen, Zenon Pylyshyn, Philip Robbins, and two anonymous referees for this journal for helpful comments and criticism. In addition, I am grateful for discussion from audiences at Auburn University and the University of British Columbia, where I presented earlier versions of this paper.

References

[Blaser et al., 2000] Blaser, E., Pylyshyn, Z. W., and Holcombe, A. O. (2000). Tracking an object through feature space. *Nature*, 408:196–199.

[Clark, 1993] Clark, A. (1993). *Sensory Qualities*. Clarendon Press, Oxford.

[Clark, 2000] Clark, A. (2000). *A Theory of Sentience*. Oxford University Press, New York.

[Clark, 2004] Clark, A. (2004). Feature placing and proto-objects. *Philosophical Psychology*.

[Jackson, 1977] Jackson, F. (1977). *Perception: A Representative Theory*. Cambridge University Press, New York.

[Johnson et al., 1991] Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, 40:1–19.

[Kaplan, 1989] Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In Almog, J., Perry, J., and Wettstein, H., editors, *Themes From Kaplan*, chapter 17, pages 481–563. Oxford University Press, New York. Originally circulated as a mimeograph from the Department of Philosophy, UCLA.

[Kremenitzer et al., 1979] Kremenitzer, J. P., Vaughan, H. G., Kurtzberg, D., and Dowling, K. (1979). Smooth-pursuit eye movements in the newborn infant. *Child Development*, 50:442–448.

[Kripke, 1980] Kripke, S. (1980). *Naming and Necessity*. Harvard University Press, Cambridge, Massachusetts.

[Matthen, 2004] Matthen, M. (2004). *Philosophical Psychology*. This issue.

[Maurer and Young, 1983] Maurer, D. and Young, R. (1983). Newborns' following of natural and distorted arrangements of facial features. *Infant Behavior and Development*, 6:127–131.

[Neisser, 1967] Neisser, U. (1967). *Cognitive Psychology*. Appletron–Century–Crofts, New York.

[Neisser, 1979] Neisser, U. (1979). The control of information pickup in selective looking. In Pick, A., editor, *Perception and its Development*. Erlbaum, Hillsdale, NJ.

[Neisser and Blecklen, 1975] Neisser, U. and Blecklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology*, 7:480–494.

[Perry, 1979] Perry, J. (1979). The problem of the essential indexical. *Noûs*, 13:3–21.

[Posner, 1978] Posner, M. I. (1978). *Chronometric Explorations of Mind*. Lawrence Erlbaum, Hillsdale, New Jersey.

[Posner et al., 1987] Posner, M. I., Inhoff, A. W., Friedrich, F. J., and Cohen, A. (1987). Isolating attentional systems: A cognitive-anatomical analysis. *Psychobiology*, 15:107–121.

[Posner and Petersen, 1990] Posner, M. I. and Petersen, S. E. (1990). The attenion system of the human brain. *Annual Review of Neuroscience*, 13:25–42.

[Posner and Rothbart, 1992] Posner, M. I. and Rothbart, M. J. (1992). Attentional mechanisms and conscious experience. In Milner, A. D. and Rugg, M. D., editors, *The Neuropsychology of Consciousness*, pages 91–112. Academic Press, London.

[Posner et al., 1980] Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109:160–174.

[Pylyshyn, 1989] Pylyshyn, Z. W. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32:65–97.

[Pylyshyn, 1998] Pylyshyn, Z. W. (1998). Visual indexes in spatial perception and mental imagery. In Wright, R., editor, *Visual Attention*, pages 215–231. Oxford University Press, New York.

[Pylyshyn, 2000] Pylyshyn, Z. W. (2000). Situating vision in the world. *Trends in Cognitive Sciences*, 4(5):197–207.

[Pylyshyn, 2001] Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition*, 80:127–158.

[Pylyshyn, 2003] Pylyshyn, Z. W. (2003). *Seeing and Visualizing: It's Not What You Think*. MIT Press, Cambridge, Massachusetts.

[Pylyshyn and Storm, 1988] Pylyshyn, Z. W. and Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, 3(3):1–19.

[Quine, 1960] Quine, W. V. O. (1960). *Word and Object*. MIT Press, Cambridge, Massachusetts.

[Scholl, 2001] Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, 80:1–46.

[Scholl et al., 2001] Scholl, B. J., Pylyshyn, Z., and Feldman, J. (2001). What is a visual object? evidence from target merging in multiple object tracking. *Cognition*, 80:159–177.

[Shore et al., 2002] Shore, D. I., Spry, E., and Spence, C. (2002). Confusing the mind by crossing the hands. *Cognitive Brain Research*, 14(1):153–163.

[Siegel, 2002] Siegel, S. (2002). Review of A Theory of Sentience. *The Philosophical Review*, 111(1):135–138.

[Strawson, 1959] Strawson, P. F. (1959). *Individuals*. Routledge, New York.

[Valdes-Sosa et al., 1998a] Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., and Pinilla, T. (1998a). Switching attention without shifting the spotlight: Object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience*, 10:137–151.

[Valdes-Sosa et al., 1998b] Valdes-Sosa, M., Cobo, A., and Pinilla, T. (1998b). Transparent motion and object-based attention. *Cognition*, 66:B13–B23.

[Yamamoto and Kitazawa, 2001] Yamamoto, S. and Kitazawa, S. (2001). Reversal of subjective temporal order due to arm crossing. *Nature Neuroscience*, 4(7):759–765.