

**Attention & Inscrutability:  
A commentary on John Campbell, *Reference and  
Consciousness*. Pacific APA, Pasadena California,  
26 March 2004**

Austen Clark  
Department of Philosophy  
103 Manchester Hall  
University of Connecticut  
Storrs, CT 06269-2054

*Philosophical Studies* 127 (January 2006): 167-193.

We assemble here in this time and place to discuss the thesis that conscious attention can provide knowledge of reference of perceptual demonstratives. I shall focus my commentary on what this claim means, and on the main argument for it found in the first five chapters of *Reference and Consciousness*. The middle term of that argument is an account of what attention does: what its job or function is. There is much that is admirable in this account, and I am confident that it will be the foundation, the launching-pad, for much future work on the subject. But in the end I will argue that Campbell's picture makes the mechanisms of attention too smart: smarter than they are, smarter than they could be. If we come to a more realistic appraisal of the skills and capacities of our sub-personal minions, the "knowledge of reference" which they yield will have to be taken down a notch or two. But first let us clarify what the argument is.

I. Attend to this

Campbell's thesis is best understood, at least initially, by confining its scope to a particular kind of situation or scene--one with particular stage-settings all in place. The scene must include at least one linguistically competent human, who is awake and alert, and who has some sensory capacities intact. It must include a linguistic token which is noticed and comprehended by that human, and which includes a perceptual demonstrative. The referent of that demonstrative must be perceptible by that human. Then, the thesis is: in such situations, or in situations relevantly like them, "knowledge of the reference of a demonstrative is provided by conscious attention to the object" (22).<sup>1</sup>

<sup>1</sup> All page references are to Campbell 2002 unless otherwise noted.

Situations that are relevantly like this dramatic scene could include those in which a human is merely thinking a demonstrative thought. Sometimes non-human animals are mentioned in the discussion, but the only relevant ones would be ones that can grasp demonstrative thoughts, and that have sensory systems and mechanisms of selective attention that work like ours do. In other words, for all we know we might as well confine the discussion to people. They must be linguistically competent people, who notice the tokening of a perceptual demonstrative, and grasp that it is a demonstrative, since failure on any of those counts would foreclose the possibility of our subject coming to "know the reference" of the demonstrative.<sup>2</sup> The referent of the demonstrative must also be perceptible. The claim: in such settings conscious attention by that subject to that object yields capacities that can accurately be described as *knowledge* of reference.

As Campbell admits in his introduction, attention and reference are terms that seem rather far apart; they are "generally taken to be different topics" (1). The argument to connect them rests on a fascinating collection of recent findings on the functional role of selective attention. Contemporary cognitive models assign to selective attention some rather surprising jobs. It has to *do* certain things to manage and coordinate an unruly hodge-podge of sensory channels; the way it goes about its job, and even the fact that the job needs to be done at all, are not intuitively obvious or apparent to introspection or common sense. The story is a surprising and interesting one, and I will focus my commentary on it. Campbell will argue that the set of capacities we assign to get the thing done can fairly be described as knowledge of reference.

Why call it knowledge? It is not knowledge *that* the demonstrative has such and such a referent; at best it is knowledge *of* the referent. It may help matters--it may hurt matters--to note that Campbell explicitly models his terminology on Russell's notion of knowledge by acquaintance. His "knowledge of reference" is "knowledge" for somewhat the same reason that knowledge by acquaintance was "knowledge". He says:

---

<sup>2</sup> Conscious attention can show up in settings in which it fails to yield knowledge of reference of a demonstrative. Perhaps the situation is devoid of any demonstrative about which to know. Or perhaps a demonstrative is uttered, but the subject simply does not hear it, or does not notice it. Or perhaps our subject is a cat, paying very careful attention to the movements of a mouse (thank you very much) but incapable of understanding a demonstrative, or of noticing that it is a demonstrative. In these and other settings conscious attention surely does things other than yield knowledge of reference.

It is a state more primitive than thought about the object, which nonetheless, by bringing the object itself into the subjective life of the thinker, makes it possible to think about that object. (6)

It is provocative, but as far as I can see strictly speaking unnecessary, to use the *K* word. It could just as well be described, simply, as acquaintance; or, even less provocatively, as a kind of primitive, direct, pre-conceptual, non-descriptive access to the perceptible objects to which one attends. Indeed there is an interesting literature within vision science about "deictic codes" (Ballard et al 1997), "visual object files" (Kahneman, Treisman, & Gibbs 1992, Wolfe & Bennett 1997), and "visual indices" (Pylyshyn 2001), all of which are postulated to yield some sort of direct and non-conceptually mediated access to visual objects of a sort similar to the one Campbell calls knowledge of reference. The terminology might be vintage 1910, but the theoretical apparatus deployed to defend it is quite contemporary.

So how does the contemporary story go? It starts with the bewildering fact that even within a "single" modality such as vision, distinct visual attributes are registered and processed in distinct visual channels. Color, motion, and form are the paradigm examples: when you see the shiny red firetruck with ladders, roaring towards you down the street, one part of your visual system manages the perception as of shiny redness, another part the perception as of firetruck-with-ladders, and a third the perception of motion thereabouts, heading this-a-way. Sound is handled by a whole separate department. So how do we manage to perceive in that scene one thing that is shiny and red and a firetruck (with ladders) and heading this-a-way? One prominent and influential answer is provided by Anne Treisman's "feature integration" model of attention, which has been the subject of more or less continuous research for over twenty years. Campbell relies on this account. According to it, the job of integrating information across feature streams so that one can perceive one thing as having multiple features is assigned, surprisingly, to the mechanisms of selective attention. Attention provides the glue that binds together features registered in distinct sensory channels when those features are perceived as features of one thing. You heard me right. *Attention* does this.

Campbell goes admirably deep into the details. We need somehow to coordinate separate bits of information about color, texture, shape, and motion. Treisman proposed a simple principle by which the coordination could be managed. At any given time, selective attention selects a particular location, and all the features found at that time to characterize that location are registered as features of one thing. To

use the current jargon, this is a "location-based" scheme for solving the "property binding" problem. It presumes that each channel keeps track of (at least) the relative locations of features it registers. Spatio-temporal location can then serve as a universally available principle by which the various stories can be collated and fit together into one story.

For this to work the system must be able to change the flow of information within itself, and do it in a way that is appropriate to the task at hand. It must have means to select the appropriate feature maps, and from them, depending on the task, select the appropriate sources and route their reports to just those other internal parties that need them, that epistemically "hunger" for them. Furthermore, this gating and re-routing of information must be instantly modifiable, and under enough control that they can be altered as needed. Selective attention is given this job: it is in charge of the gates and switches, the filters and pathways, with which to change and control the flow of information within the system. By attending to a stimulus one changes the disposition of one's own information processing resources with regards to that stimulus. Specifically, according to Treisman, by attending to a location one opens the channels relevant to (and closes the ones irrelevant to) collecting all the information one's sensory systems have registered about the location in question. Discriminations of features at that location become faster and more reliable; those at other locations become slower and more error-prone. Selection by selective attention has quite real effects. Recently, the search for neural correlates in this domain has become, as Nancy Kanwisher (2001, 89) put it, "*wildly* successful". There are multiple confirmations that neurons in the specific cortical areas subserving discriminations of the features in question increase their metabolism and blood flow; areas not so activated by the selective attention decrease their activity. By focusing your attention you change the blood flow in your head. You change the metabolism rates, the thresholds, and the baseline activation rates of neurons in different parts of your nervous system. And you do this at will, effortlessly, by simply shifting your attention, or allowing it to drift. Subjects are presented a rivalrous stimulus, with a picture of a face superimposed on a picture of a house. Neurons in the fusiform face area (FFA) become more active when the subject is paying attention to the face component; neurons in the parahippocampal place area (PPA—a shape perception area) become more active when attention shifts to the house (see Kanwisher 2001, 92).

One way to summarize these interactions is to say that selective

attention "is what causes and justifies the use of information-processing routines in verifying propositions about that thing." (28) A mouthful, but an accurate one! The information processing routines in question are all the specific feature maps and modules activated when selective attention selects the thing in question. They provide all the resources the system has available to verify or justify or test implications of propositions about the thing in question. Furthermore, activation by selective attention causes those resources to become available: unless there is such activation, the needed gates and pathways are simply not open, and central resources are devoted elsewhere. So it seems fair to say that, in situations of the sort in question,

1. If one directs conscious attention to the referent of the demonstrative, then one has means that cause and justify the use of particular information processing procedures to verify and find implications of propositions about the object in question.

To this we need add only a second main premise:

2. In such situations, that which causes and justifies the use of particular information processing procedures to verify and find implications of propositions about the referent of a demonstrative is worthy of the label "knowledge of reference of the demonstrative".

And with that the main conclusion would be secured: conscious attention can provide knowledge of reference of a demonstrative.

This is a somewhat weak reading of Campbell's main argument, though it has the great virtues of being both valid and (I think) defensible. Both premises are cast simply as conditionals, not biconditionals. The conclusion states a sufficient condition: In some situations, conscious attention can suffice for *K*. Campbell in places suggests stronger readings of both of the premises and of the conclusion, though I don't see how any stronger version of the argument can be made to work.

So for example consider premise (1). One caveat about premise (1): there is a distinction, ignored thus far, between "selective attention" and "conscious attention". The former is the term used in the experimental literature. The latter is Campbell's. The problem is that the two are not necessarily equivalent. Something might do all the binding, selection, gating, and information-processing jobs attributed above to "selective attention" even though one is not aware of the

object thereby bound. Treisman (1998, 1303) reports on "negative priming" experiments which she takes to demonstrate the possibility of "implicit binding". The latter would be binding without awareness of the object bound. It may not even be true that the psychologists' notion of "selective attention" warrants the inference that one is conscious of that to which one attends.

Second, Campbell in some places seems to suggest that not only is (1) true, but the converse of (1) is as well. That is, conscious attention is (perhaps) *necessary* to cause and justify the use of particular information processing procedures to verify and find implications of propositions about the object in question. (See pp. 25, 27). I think the suggestion should be resisted. A moment's reflection can suggest several ways in which, for example, propositions about the referent of a demonstrative can be verified without the verifiers ever having the opportunity to focus attention upon that object. They might do so without even perceiving it. (We are FBI agents with a wiretap on the bad guys. We hear one say "that's the gun he used to whack Hoffa", and the other says "hand it to me".<sup>1</sup> Later the lab finds both their fingerprints on exactly one of the guns. It seems we have determined which one they were talking about, even though we never perceived the gun ourselves, much less attended to it.)

Campbell's argument for the second premise (in ch 2) is based on classic accounts of what it is to know the reference of a term, cast in terms of introduction rules and elimination rules for the term in question. That is, what must one know to be able successfully to introduce a demonstrative term into some context of discourse? Likewise, how can inferences which eliminate it be justified? This discussion suggests that (2), unlike (1), could plausibly be strengthened to a biconditional. One can read that discussion as providing an account of what knowledge of reference *is*: what constitutes it and what it requires. I have no strong objections to this reading. As already noted, it is something of a term of art to call this "knowledge" at all, and one is free to use a term of art however one pleases.

---

<sup>1</sup> In case some readers don't follow news of American gangsters: James Hoffa, ex-president of the Teamster's Union and father of the current president of that union, disappeared on 30 July 1975 while traveling to meet a reputed gangster in Detroit. His body has never been found. I hope the meanings of "snitch", "wise guy", and "whack" are clear from the context.

## II. Talk to Sally

The more compelling arguments for premise two arise from answers to a slightly different question that Campbell poses in the beginning of the book:

How exactly does your identification of an object, at the level of your subjective life, bear on the selection of information for further processing? That is, what is it about your identification of the object at the level of your subjective life that causes the selection of just the right underlying information to control your verbal reports? (17)

Quite apart from the spooky physiological effects of selective attention—what seems to be the do-it-yourself re-wiring of your own nervous system, performed on the fly, adventitiously, without forethought or anesthetic—the other wondrous aspect of this capacity of ours is that selective attention is, to some degree at least, under voluntary control, and it is sensitive to directions formulated in sentences. In a search task, for example, subjects might be told to hunt for a letter "T" that is red, and if they find it press a key as quickly as possible. These instructions then enable the subject to deploy the appropriate visual cognitive resources so that, somehow, the task can be managed. Channels to the parts of the visual nervous system that subserve the discrimination of colors and shapes are opened, primed, and facilitated; other irrelevant visual features, as well as all the other non-visual sensory channels, are gated or suppressed. Somehow too the word comes down from on-high that it is RED that is sought; the word "red" starts a chain of processes that eventuate in the appropriate instruction reaching chromatic systems in such a way that the appropriate chromatic target is identified. How do you communicate with your sub-personal feature-processors so as to enable them to do this? How do you even know which parts of the nervous system to address when issuing these instructions?

At some point non-conceptual representation must, as it were, talk to the conceptual variety. Search tasks requiring directed attention provide a forum within which, it seems, such an interchange—such "talk"—must happen. Campbell asks the bold and intriguing question: *how* do they talk to one another? He argues there must be a "commensurability" between them—between what he calls the level of "conscious attention" and the underlying "information-processing":

The issue that concerns me is, as it were, a top-down commensurability: to explain in detail just how conscious attention to an object can identify the thing so that, at the information processing level, just the right information is selected to control your verbal reports. How in

detail does conscious attention to an object serve to single out the right information to control verbal report and action? (17)

We have seen that selective attention employs "binding principles"—in particular, location—to manage the collating of features across different input streams. A "binding principle" or "parameter" is defined as "the characteristic of the object that the visual system treats as distinctive of that object, and uses in binding together features as features of that thing." (37) So binding principles are defined as principles that operate at the information-processing level; they are postulated to explain how commerce can proceed across distinct feature maps. If color is processed separately from shape, then we face a puzzle of engineering to connect them together; location as a binding principle is a potential answer to that question.

Campbell suggests that the same binding principle might also serve as the medium providing commensurability between the levels of conscious attention and those feature maps. Per hypothesis, attention to a location serves to bind together features sensed at that location at the time of the attending. The location parameter is potent at organizing information from multiple feature maps, and (per hypothesis) it is also a parameter accessible to attention. Differences in location define the differences in that to which one can attend (per hypothesis). Campbell's suggestion is elegant: use that same binding parameter as the medium of interchange between "conscious attention" and "information processing":

conscious attention to an object has to be able to identify that object for the benefit of information-processing systems. And the natural way for conscious attention to identify the object, for the benefit of the information-processing systems, is to use the parameter that was used in solving the Binding Problem for that object... whatever complex parameter we use in solving the Binding Problem, that will provide a kind of address for the object that is bound. And the way for conscious attention to identify the object, for the benefit of the information-processing systems, will be to use that complex parameter to identify the thing. (41)

The key idea is that "experienced location" serves as the principle providing commensurability: it provides something like an address, Campbell says, at which the bundle of features can be found. It also provides an address that can be handed over the "visuomotor routines" to identify the target of actions. (The connection to visual action is important, and gives some independent grounds for thinking there must be "commensurability" of some sort between these systems. But I will leave those details aside.)

For this to work, there must be some intrinsic aspects of your visual experience of the object, which can identify, for the benefit of subsequent information-processing, either to verify or to initiate action, which object is in question. But how exactly does your awareness of the object identify the target? ...I am suggesting that a central aspect here is the experienced location of the object; or, more generally, the complex parameter used in binding. (42-43)

This he thinks can explain why experienced location is important to the function of perceptual demonstratives, even though it does not pick out its referent by description. The experienced location does not provide a description of the location, but it does contribute to the sense of the demonstrative. Its use as a binding principle gives experienced location a special role, which is difficult to characterize without some characterization of how binding proceeds. It can, says Campbell, contribute to the sense of the demonstrative, but not by adding conditions to a description used to identify the referent. The special role assigned to experienced location distinguishes Campbell's account, he says, from those of both Kaplan and Evans.

The rather complicated model is made clear by a charming analogy. Suppose you belong to a social group in which people gossip about one another. Philosophers of course *never* gossip about one another, and they never, *ever*, worry about reputation, prestige, or social standing; but let us suppose you occasionally associate with lesser humans. In particular, you have an informant, named Sally. Campbell says:

Suppose there are several different people called Reagan, and your informant Sally collects and provides you with information about each of them. So you stand to Sally somewhat as the level of conscious attention stands to the level of visual information-processing. Or rather...the level of conscious attention stands to the level of information-processing somewhat as the content of your speech stands to the level of Sally's speech, in the analogy. Now suppose...

- (a) You want to be able to interrogate Sally for further information about any of the people about whom she is giving you information, and
- (b) You want to be able to instruct Sally to act on any one of the people about whom she is giving you information.

To have these general capacities, you have to be able to identify the person in whom you are interested, for Sally's benefit. That is, your identification of the person about whom you wish to interrogate Sally has to be one that she can use to find the further information you have requested. Not just any way of uniquely identifying the thing you have in mind will do. (39)

You must be sure she understands you when you interrogate her or give her orders. Suppose for example we try to communicate with Sally by identifying Reagan as "the famous actor who became governor of California". (No, not *that* actor who became governor, the other one. Campbell must be pleased at the way this example has improved with age!) The difficulty is that there is no way to assure ourselves that Sally understands any such descriptions. A simple alternative would be to use the same tokens that Sally uses when she passes along her gossip. If she uses index cards, one could write queries or instructions on those same cards. Campbell suggests that something like this could work as well for ensuring a commensurability between conscious attention, visual information processing, and visuomotor action.

To spell out the parallel: when one attends to something, one experiences it as having a location. Experienced location provides something like an address that can be used to identify objects at that address. It turns out, somewhat surprisingly, that such addresses are essential to the success of property binding; features are bound together because they are at the same location. So experienced locations have a side to them that already talks quite fluently to the underlying feature-processing mechanisms. Furthermore, the same sort of address might work to identify a target for visuomotor routines. In short, these addresses provide a perfect means of informational interchange among these systems. Using them, we can figure out whom Sally is talking about, we can ask her to find out something about someone else, and we can tell her to do things to people. (Sally: who is that wise guy? Sally: find the snitch. Sally: whack that guy!)

### III. Be appropriately stylish

Things are fine up to this point. I am entirely sympathetic with developments thus far. There are of course nagging empirical questions and theoretical problems with Treisman's model, with the account of property binding, and with the very strong emphasis on location-based models of selection. Recent evidence shows that so called "object-based" selection might go much deeper (or start earlier) than Treisman et. al thought. But basically, if location-based accounts turn out to be entirely wrong, then *both* Professor Campbell and I are wrong. We're in the same boat, paddling the same canoe, up to this point in the navigable waterways.

*At* this point we have endowed the attention/perception interface with a preconceptual and direct means of picking out locations; the things sensed are identified merely as that which is found at that

location. There is no descriptive backing, no wielding of criteria of individuation (beyond the meager ones granted by discriminabilities of location), no criteria by which "the same one again" could be distinguished from "a different one, but qualitatively identical". But Campbell goes on to argue that study of the innards of the attentional processes can yield a richer harvest. They give one, he thinks, a rather full answer to the question "to which object are you attending?", one which is capable of settling questions such as "is it the same one as the one to which you were attending a moment ago?", and "is it the same as the parcel of matter composing the thing at that place?" and "is it distinct from a series of temporal stages?". Facts about the innards of attention can settle questions that one might think—and that Wiggins and Quine and Davidson have argued—require mastery of sortal concepts and of an apparatus of individuation. I don't think our little canoe can make it through these rougher waters. I am afraid I must abandon ship, and leave Campbell to paddle through those rapids alone.

The problems start with what might appear to be a small extension of the notion of "binding parameters" to what he calls "complex binding parameters" or "styles of conscious attention". Campbell says:

Different styles of conscious attention will be used in attending to different sorts of object. For example, if you are consciously attending to a person over a period of time, the way in which you keep track of that person will be quite different from the way in which you keep track of a valley to which you are attending over a period of time. ... These differences in style of attention amount to differences in what I called the complex binding parameter used by the visual system in putting together the information true of the object. The binding parameter for a person will have to allow for the possibility of movement by the person; the binding parameter for a valley will not... So the style of conscious attention to the object that is appropriate will depend on what sort of object is in question. (62)

These "complex binding parameters" or "styles of conscious attention" are postulated to do what is necessary to manage and collect information from the same object over time or over multiple presentations. By the end of the discussion they have been promoted into theoretical entities with powers at least equivalent to those of an apparatus of sortal concepts:

Our grasp of the identity conditions of an object over time, or the boundaries of the object at a time, is grounded not in grasp of sortal concepts, but in the style of conscious attention that we pay to a thing. And conscious attention to the object does not have to be focused by a grasp of sortal concepts; the various styles of conscious attention of

which we are capable do not rely on our use of sortal concepts. (83)

Campbell mentions Wiggins' example: how do we distinguish between the river, the parcel of matter making it up, and fusion of the components of that parcel? This is not particularly hard, he says; "we can answer this question by appealing to what I earlier called a complex binding parameter as focusing the subject's conscious attention": (69-70)

the singling out of an object in experience need not involve the application of sortal concepts; only the mechanisms of binding. Whether you are consciously attending to a river or a mass of molecules, for example, will show up in how your visual system binds together information from the thing over time. If you have to keep moving downstream to keep track of the object of your attention, then you are attending to a collection of water molecules rather than a river. If, on the other hand, you are binding together information from any point in the course of the river, as all relating to a single object, then you are attending to the river itself. The distinction between consciously attending to a collection of water molecules, and consciously attending to a river, is not particularly hard to draw, even without appealing to any grasp of sortal concepts by the subject. (70)

Likewise, Quine's claim that we need here to invoke an "apparatus of individuation" including sortal concepts is wrong, for similar reasons:

According to Quine, the involvement of the sortal concept is needed for there to be a determinate answer to the question 'To which object is the subject consciously attending?'. If we do not appeal to the subject's grasp of a sortal concept, how could we say what the difference is between attending to a river and attending to a collection of water molecules, for example? I think we can answer this question by appealing to what I earlier called a complex binding parameter as focusing the subject's conscious attention... we should think of the complex binding parameter used as playing a role also in conscious attention to the object: it provides, in effect, a way of identifying the object which can be used to find information about that object in various processing streams. (69-70)

The discussion is cast in terms of what Campbell calls the "Delineation thesis". This is a thesis about how attention is focused: "Conscious attention to an object has to be focused by the use of a sortal concept which delineates the boundaries of the object to which you are attending." (69) Campbell's analysis of this thesis seems on-target. If sortal dependence entailed the Delineation thesis then sortal dependence would be a silly doctrine. A hungry carnivore could have its attention focused upon *you* even though it lacks sortal concepts that delineate your boundaries. Skeptics who travel are strongly advised to

avoid the savanna, parts of rural India, and the high country in this very State, where joggers have been killed by mountain lions.

But it seems a bit unfair to treat sortal dependence as a claim about how to focus one's attention. It sounds distinctly *odd* to read Quine as asking the question 'To which object is the subject consciously attending?' and then implicating sortal concepts as a necessary part of the answer (69). The issue as discussed by Quine and Wiggins and Davidson is rather: if we think it true that in using some expression a subject is referring to object  $x$ , and not to  $y$ , what are the conditions necessary for this to happen? To answer that question it seems we must consider how the subject represents  $x$ , whether the subject represents the thing as an  $x$ , or as a  $y$ : and there the suggestion that mastery of sortal concepts might be required is not so silly at all.

Clearly there are visual processes of segmentation and grouping that in some sense serve to "pick out" or "highlight" something visually. These principles tend to be satisfied by anything that is seen *as* an object. They include the ones Eli Hirsch listed a while ago: distinct boundaries (or, where there are no distinct boundaries, jointedness at convexities); qualitative homogeneity; symmetrical shape; separability; dynamic cohesiveness (Hirsch 1982, 105ff). The "things" selected by visual selective attention are "delineated" by principles such as these. Vision "singles out" a thing insofar as "visual objects" (things seen as objects) must satisfy these principles. So up to a point it is fair to talk about visual attention "singling out" a thing. As Campbell argues, contrary to the Delineation thesis, such delineation does not require sortal concepts.

However, as far as "singling out" objects goes, these grouping and segmentation principles exhaust the capacities of the visual system. If we have two distinct sorts of objects that equally well satisfy all these perceptual principles, vision will be indifferent between them. It can't tell the difference; either serves equally well. If we want a fair test of sortal dependence, then, the alternatives should include some candidates that in this way pass all the tests that the visual system imposes. In particular, can we imagine subjects whose visual systems work just like ours do, but who routinely and systematically interpret visual demonstratives in a different way?

These individuals have just the visual experiences that we do—their visual systems work like ours, what is bound in us is bound in them, and so on—but when confronted with a "that" whose referent is to be assigned visually, these individuals understand the demonstratives to refer to a wholly different class of objects than we do.

Let us confine the discussion to perceptual demonstratives whose

referent is to be assigned visually—what Campbell calls "visual demonstratives". So if I can point at something perceptible only visually, say "that is a Great Egret", and thereby say something true, the token of "that" is a visual demonstrative, and (in Campbell terminology) my conscious visual attention gives me knowledge of its reference. It picks out a bird. Furthermore, different "styles" of conscious attention give me knowledge of different varieties of spatiotemporal continuity, different identity conditions and persistence conditions appropriate to different kinds of objects. Binding principles within vision do this; sortal concepts are not necessary.

I think it is not at all difficult to construct a class of inscrutably different objects to serve as referents for these demonstratives. They are inscrutably different, that is, under the full scrutiny of the visual system. They equally well satisfy all the facts about binding principles and the innards of visual attention. The key to the construction is to remember that the visual system is, well, *visual*; visually equivalent things will be, to it, equivalent.

Here is a principle of visual equivalence. For any  $x$ , for any person  $p$ , for any time  $t$ ,  $x$  is visible to  $p$  at  $t$  if and only if some portion of  $x$  is visible to  $p$  at  $t$ . Call the portion of  $x$  that is visible at that time the "visible portion" of  $x$ . It might in rare cases include the entirety of  $x$ — $x$  would have to be a translucent object in full view—but most of the time we can't see the backsides of things, and often even their facing surfaces are occluded to a greater or lesser degree by other things. Part of the Great Egret is hidden by the tree, and I don't actually see its right foot; but there is a portion that I do see.

Our first class of benighted ontologists take visual demonstratives to refer not to things, but to the visible portions of things. We take "That is a Great Egret" to mean "That = an  $x$  such that  $x$  is a Great Egret". They take it to mean "That = a visible portion of some  $x$  such that  $x$  is a Great Egret." We think we see things; they insist that strictly speaking we (usually) see only portions of things. They decry our loose way of speaking; for them a "thing" is not something you see (heavens no!), but is instead a theoretical entity, postulated or constructed to explain our somewhat haphazard histories of sightings of visible portions of things. Because the referents of demonstratives are merely portions of things, let us call these people "portionalists". They resemble real figures in the history of philosophy; I think both Russell and Moore *were* portionalists at points in their careers.

The next step is to imagine that you swap visual careers with a portionalist. That ghostly track of points in space *from* which you have seen all the things you have ever seen is occupied, instead, by a

portionalist. More simply, the portionalist lives through exactly the glimpses and sightings and all the visual experiences that you have had. What's odd about this, of course, is that from the eyeball's point of view, nothing changes. The portionalist's visual system would have exactly the visual inputs that your visual system had. Every single bit of the terabits per second of visual information that was input to your visual system would, likewise, be input to the visual system of the portionalist.

The task of managing the information in one system is identical, therefore, to the task of managing the information in the other one. If one system uses property binding to help manage that information, then the other could too. If locations serve as a binding parameter in one, they could also in the other. The information processing tasks are identical, so any stratagem useful in one of the systems would be equally useful in the other.

Finally, imagine that the psychologists at your home campus propose to do a series of experiments on visual selective attention. You volunteer as a subject, and complete the experiments with your typical panache, manifesting those operating parameters *K*, whatever they are, that characterize your visual selective attention. I suggest your portionalist doppelganger could likewise volunteer, and likewise sport precisely the parameters *K*. Remember, any stratagem your visual system uses to manage visual information would be duplicated in the portionalist. If there were portionalists mixed in our subject pools, we could not discern their presence from the data. Experiments on visual selective attention would have exactly the same results whether done on us or done on portionalists. So none of the facts about visual selective attention could differentiate us from portionalists. So none of the facts about visual selective attention preclude the portionalist understanding of the reference of visual demonstratives.

How is it precluded, if at all? I think Quine and Wiggins and Davidson are right: it is only precluded if one can engage in rather sophisticated discourse with the hapless subjects. Portionalists understand "thing", "portion of thing", "same thing", and so on in different ways than we do, and the differences are manifest only if we assume a fixed interpretation for at least some pronouns, plurals, and the identity sign, and can then ask careful questions about persistence and individuation. "Is this one part of that one? Is that still the same one? Do you and I both see it?" and so on. Our interlocutor must wield sortal concepts, identity, pronouns, and plurals for this to be possible.

Portionalists think that visual demonstratives do not refer to what we think of as physical things, but to some other entity, distinct from a

physical thing, that (a) is directly visible, and (b) stands in some "peculiarly intimate relation, yet to be determined" to physical things.<sup>1</sup> As long as one takes care to duplicate all the inputs exactly (all the "hard facts", all the "facts of sensible appearance"), this schema can be used to generate a family of views, distinguished from one another by the kinds of entities proposed. As I mentioned, some of the offspring in this family were sired by real philosophers. Two dimensional color patches in the visual field and the "facing surfaces" of things provide two familiar familial examples. Sense-data are perhaps the black sheep of the family. But other offspring are good citizens, and can live in complete harmony with everything we know about the innards of visual attention.

Perhaps, for example, visual demonstratives refer to visual solid angles. A "visual solid angle" is a notion dating back to Ptolemy: it is a three dimensional region of visual perimetry, described in retinotopic terms (see Gibson 1979, 68-69). The point at which you are fixated is at azimuth 0, altitude 0. Imagine tracing an outline around just that portion of the Great Egret that is visible. Each point on that outline likewise yields an azimuth and altitude. Those coordinates specify directions, and one can think of each one as a ray or a vector, proceeding outwards from the eye, in precisely that direction. They go as far as the eye can see. Collect the series of them around the entire outline, and you get a three dimensional cone. That cone is a "visual solid angle". It is a somewhat weird three dimensional thing. Its outer envelope intersects the outline of the bird precisely.

Deviant ontology number two treats visual demonstratives as referring to visual solid angles: the ones that enclose the visible portions of the demonstrated thing. In honor of their inventor, I will call these ontologists "Ptolemaists". They understand "that's a Great Egret" to mean "that = a visual solid angle within which there is a portion of a Great Egret". "You see it too" means you too have a visual solid angle which intersects that portion. The Ptolemaists disagree with portionalists. "That" by no means refers to portions of things. Heavens, no; the portionalists are speaking loosely. Instead it refers to the visual solid angles *enclosing* those portions. Visual equivalence is still preserved, since:

---

1 "Whenever I truly judge that *x* appears to me to have the sensible quality *q*, what happens is that I am directly aware of a certain object *y*, which (a) really does have the quality *q*, and (b) which stands in some peculiarly intimate relation, yet to be determined, to *x*." C. D. Broad, *Scientific Thought* (1927, 239).

- (a) you see  $x$  if and only if you see a visible portion of  $x$ , and  
 (b) you see a visible portion of  $x$  if and only if in your ambient optic array there exists a visual solid angle enclosing a portion of  $x$ .<sup>1</sup>

This is not entirely nuts; I am not being *entirely* silly. A fundamental task of early vision is to "segment" the ambient optic array: partition it into non-overlapping regions, whose local features can be analyzed in some independence from one another. Segmentation has robust psychological reality (see Nakayama, He, & Shimojo 1995), and visual solid angles are one way to define the segments. There almost certainly are segmentation and grouping processes that operate so that (for example) points within the Great Egret outline are grouped together and thereafter treated *as* members of a group. They stand together *as* a figure against a background. Features therein are features *of* an identified segment. If early vision has grouping processes of this sort—which it does—then the groups that are formed could provide values for visual demonstratives.<sup>2</sup>

A small modification to visual solid angles yields a third such alternative. The visual demonstrative refers, not to the full three dimensional visual solid angle, but rather to the sum of the points within it that are visible. Each such point will be at some depth from the viewer. Their sum is often a facing surface of a thing, but not always. Clouds in the sky and mists in the valley do not have surfaces. These sums of points are not strictly two dimensional entities, since different points within such a sum can have different depths. If it is a surface of an object, that surface can have perceptible "tilt" and "slant" angles. But the sum of points is not strictly a three dimensional entity, since at a given azimuth and altitude we need only one depth. Marr called these entities two-and-a-half dimensional; they are values for the 2.5 d sketch. Because they are distinct from surfaces, and they have an extra fractional dimension, I will call these entities "superficies". (A thing likened to a surface; the outward form or aspect. L16-L18. OED.) The ontology is superficialism. Models of visual surface representation employ variables that range over these

entities.<sup>3</sup> Perhaps, says the superficialist, they could also serve as values for the visual demonstratives. If we are confronting the facing surface of an object, the interpretation is familiar: "That's a Great Egret" then means "That = a surface of a Great Egret." But superficieses generalize. "That is cirrus" means "That superficies is cirrus." Or: that is an outward form, and it has the cirrus features.

Of course, like all the alternatives, deviant ontology number three requires compensating adjustments elsewhere in how one understands the lingo. But those linguistic adjustments are all made outside the purview of the visual system. Within the visual system, nothing would be perturbed; all would be exactly as it was. If our case is to be made solely on the basis of facts about the operating parameters of the visual system, I don't see how we could rule out this interpretation of the visual demonstratives. To vision the switch would be inscrutable.

#### IV. Rein in the talk talk

There is something to the talk of visual attention "singling out" a thing, and I have suggested that it should be unpacked by detailing the principles of visual segregation and grouping, by which a figure is seen as a figure, distinct from the background. Descendants of the Gestalt principles may play a role here. There are also principles within the visual system that things tend to obey if they are to be perceptible by us *as* objects: principles such as boundedness, articulation at convexities, symmetries of shape, separability, cohesive movability. But once we have finished listing such principles, we have basically finished the list of requirements known to be imposed by the visual system if it is to "pick out" something *as* a visual object. And from the point of view of ontology, that list is not particularly discriminating. Any two systems of objects that satisfy those principles will be indiscriminable to the system; it is blind to their differences. Visible things and visible portions of things provide one such pair. It is, after all, a *visual* system. If these things look exactly the same, then, as far as vision can tell, they are the same.

My guess is that Campbell has simply over-generalized the notion of "binding parameters", granting them powers that no vision scientist would endorse. This is not entirely his fault; the term "binding" has

1 Your ambient optic array (at a given time) is here presumed to be the ambient array of all the  $x$  such that you see  $x$  at that time. Gibson would consider this to be an ambient optic array whose "point of observation" is occupied by you. An alternative would be to construe "visual solid angles" to imply that their occupants are seen.

2 Segmentation and surface processing may well use something like the envelopes of visual solid angles and superficieses. I think the ontology of vision is still an open question; the answers are not yet known.

3 A paradigm example is a bounded surface located at a perceptible depth, angle, and tilt, adjoining or occluding or occluded by other such surfaces; and typically with additional perceptible features of color, gloss, lighting, shading, and texture. Computer graphics programs manipulate similar entities (typically triangles, since three points define one plane, and it is handy to be able to represent slant with just one normal vector).

been applied to all sorts of different problems of vastly different orders of complexity. Essentially any phenomenon involving any kind of grouping can be, has been, or will be called a binding problem. But notice just how powerful "complex binding parameters" have become by the end of the discussion. Campbell says:

There are many varieties of spatiotemporal continuity, appropriate variously to dogs, stars, and rivers. But the point is that the visual system can be using the appropriate type of spatiotemporal continuity, in keeping track of the thing, whichever it is, from moment to moment, and in binding together the properties of the objects as all properties of a single thing. There is no evident reason why there should be any involvement of the subject's grasp of sortal concepts in this... (82)

He says that given adequate understanding of the mechanisms of binding, it is "not particularly hard" to distinguish between "attending to a river" and "attending to a collection of molecules". Some work *has* been done on the distinction between location-based and object-based selection for selective attention, and there are various kinds of "binding problem" discussed, but as far as I know the notion that there exists one kind of complex binding parameter which is used for people, another for valleys, a third for rivers, and so on, is new, and unique to Campbell. A "complex" binding parameter must be sensitive to a particular *variety* of spatiotemporal continuity. It already includes the correct kind of persistence conditions and individuation conditions for the object in question, be it person, valley, dog, star, river, parcel of matter, or collection of molecules. These complex fellows solve problems that are immensely more sophisticated than the simple puzzle of property binding. In particular, property binding requires *no* sensitivity to persistence over time; no individuation; no numerical identity.

Campbell is correct, I think, in the following claim: if complex binding parameters could do the job he specifies, then there would be no need to invoke the subject's mastery of sortal concepts. The reason, though, is that these "complex binding parameters" already have the expressive power of sortal concepts; they already have built into them all the sensitivity to persistence, individuation and identity normally thought to be provided only by mastery of an apparatus of sortal concepts. Consider again his description. There are different *styles* of conscious attention; different *varieties* of spatiotemporal continuity. The visual system must first determine what *sort* of object it confronts. Then it can apply the *appropriate type* of binding parameter to objects of that *sort*. In everything but its name, this sounds like a sortal. It

almost sounds as if a sub-personal processor, endowed with mastery of complex binding parameters, could go toe-to-toe with the entirety of David Wiggins, matching all the distinctions that Wiggins can make using full predicate logic with identity and an apparatus of sortal concepts. Surely they're not quite *that* smart!

There is a tremendous intellectual challenge that we all face, if we are to navigate in this problem-space. We have to describe the systems of communication between sub-personal parts without over-intellectualizing; without packing more content into those signals than is warranted, or making the minions at either end of the channel smarter than they could be. I'm afraid that Campbell has done just this. Recall the analogy of talking to Sally. You are conscious attention; she is your informant and intermediary, who carries out your commands and tells you what is going on down in the information-processing sweat-shops. Clearly there must be some information interchange between conscious attention and the information-processing minions. But the risk in the analogy of talking to Sally is that we will assume it is something like *talking*. In particular, if study of the codes interchanged between attention and feature-processors really could solve the problems of inscrutability and sortal dependence raised by Quine and Wiggins, then talking to Sally really *would be* talking. A subpersonal part of you would have to be able to say to another subpersonal part something on the order of: no, no Sally; not that parcel of matter: the river! the river itself! Or: aim for the whole rabbit, not just the current stage of the rabbit! It might be delightful to discover and listen in on such intelligent patois between our subpersonal parts, but I doubt it will happen.

What will happen? A more conventional, a less bold view, goes like this. There are at least three distinguishably different levels of representation within the visual system; three different ways of representing the things that eventually (at level three) we represent *as* things. The first (most primitive, and earliest) is the "feature" level: the representation of sensibly discriminable features at spatially discriminable locations. I call this feature-placing. It registers the information necessary (for example) to *detect* an edge. The principles that operate to segregate figures from ground must operate on something: they operate on clusters of features, found on this level. Location as a binding principle makes perfect sense when understood as a way of coordinating the information in distinct feature maps. They indicate the incidence of features at places, and do this with a scheme of representation that neither requires nor warrants description using terms of individuation, identity, and reidentification. Such

processes "pick out" things insofar as they differentiate locations, but even to claim that they "identify" things is, I think, a stretch.

Once figure-ground segregation and grouping principles have been applied to what is initially nothing more than features-at-places, we allow entry into a second level: the "mid-level object-based attention system" (see Carey & Xu 2001, 181). Here we have "visible objects", e.g. things seen *as* objects. Such things are pledged to obey the Gestalt laws, if there are any; more generally they satisfy the principles of visual grouping and segregation. They are the "objects" tracked in multiple object tracking; they are the entities indexed by visual indices; they are the things that seem to move in apparent motion.

Third is a conceptual level of representation, allowing "kind-based object individuation". Here decisions about individuation and numerical identity are based on an apparatus of sortal concepts. This is the level at which Wiggins-style distinctions can be made; with it we can distinguish between the portion of the river in front of us and the parcel of matter within that portion.

The conservative and timid view (my view) is that talk of "binding principles" makes sense (it has determinate empirical application) primarily (and perhaps only) at level one: features. Certainly location-based binding only makes sense there. There are additional principles by which the visual system aggregates clusters of features and segregates them from the background, and these principles yield the "visible objects" that are tracked by the mid-level object-based attention system. But those principles are not particularly discriminating, and objects from rival ontologies (such as things *v.* visible portions of things) can equally well satisfy all of them. So even the mid-level system cannot make the sortal distinctions noted by Wiggins and Quine.

Campbell, in contrast, argues that binding principles, born on level one, can yield complex varieties that, in effect, do it all. Studies of styles of conscious attention can make all the distinctions that the conventional view makes only on level three. Whatever else one says about it, this is a tremendously bold proposal, and I should say that I very much admire the boldness of the line that Campbell takes through these largely uncharted and unknown waters. I have my doubts whether the claim actually works, but perhaps that is simply because I lack nerve. Remember I abandoned ship long ago. I got out of the canoe when I first heard the rapids coming.

## References

- Ballard, Dana, Hayhoe, M. M., Pook, P. K. & Rao, R. P. N. (1997). Deictic codes for the embodiment of cognition. *Behavioral & Brain Sciences* 20 (4): 723-67.
- Broad, C. D. (1927). *Scientific Thought*. London: Routledge and Kegan Paul.
- Campbell, John (2002). *Reference and Consciousness*. Oxford: Clarendon Press.
- Carey, Susan and Xu, Fei (2001). Infants' knowledge of objects: beyond object files and object tracking. *Cognition* 80: 179-213.
- Clark, Austen (2000). *A Theory of Sentience*. Oxford: Oxford University Press.
- Gibson, James J. (1979). *The Ecological Approach to Perception*. Boston: Houghton Mifflin.
- Hirsch, Eli. (1982). *The Concept of Identity*. New York: Oxford University Press.
- Kahneman, D., Treisman, A. & Gibbs, B. J. (1992) The reviewing of object files: object specific interpretation of information. *Cognitive Psychology* 24 (2): 175-219.
- Kanwisher, Nancy (2001). Neural events and perceptual awareness. *Cognition* 79: 89-113.
- Nakayama, Ken, Zijiang, J. He, & Shimojo, Shinsuke (1995). Visual surface representation: A critical link between lower-level and higher-level vision. In Stephen M. Kosslyn & Daniel N. Osherson (eds). *Visual Cognition. (Invitation to Cognitive Science, 2nd. ed., vol. 2. General editor Daniel N. Osherson)*. Cambridge, Mass., MIT Press, 1-70.
- Pylyshyn, Zenon (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition* 80: 127-158.
- Treisman, Anne (1998). Feature binding, attention and object perception. *Phil Trans R. Soc London B* 353: 1295-1306.
- Wolfe, J. M. and Bennett, Sara C. (1997). Preattentive object files: shapeless bundles of basic features. *Vision Research* 37 (1): 25-43.