

Concept empiricism and the vehicles of thought

Daniel A. Weiskopf

Abstract: Concept empiricists are committed to the claim that the vehicles of thought are re-activated perceptual representations. Evidence for empiricism comes from a range of neuroscientific studies showing that perceptual regions of the brain are employed during cognitive tasks such as categorization and inference. I examine the extant neuroscientific evidence and argue that it falls short of establishing this core empiricist claim. During conceptual tasks, the causal structure of the brain produces widespread activity in both perceptual and non-perceptual systems. I lay out several conditions on what is required for a neural state to be a realizer of the functional role played by concepts, and argue that no subset of this activity can be singled out as the unique neural vehicle of conceptual thought. Finally, I suggest that, while the strongest form of empiricism is probably false, the evidence is consistent with several weaker forms of empiricism.

1. Introduction

Many empiricist theses fell on hard times in the latter half of the 20th century. As a semantic thesis, empiricism fell to Quine's critique of verificationist conceptions of meaning. As a thesis about the mind's innate contents and mechanisms, empiricism was a primary target of Chomsky's arguments concerning language learning and the poverty of the stimulus. Finally, empiricism suffered from its traditional connection with purely associationist models of mental processing, which lost ground to computational models implementing rule-based transitions among logically structured representations (at least until the connectionist revival of the late 1980's).

Despite this series of attacks on a variety of empiricist claims, one form of empiricism, namely *concept* empiricism, has been enjoying a resurgence in recent years among philosophers and psychologists (Barsalou, 1999; Goldstone & Barsalou, 1998; Prinz, 2002). Contemporary concept empiricism can be defined as a thesis about the nature of the vehicles of thought. This *perceptual vehicles thesis* states that thoughts are composed of perceptual representations, or

copies thereof.¹ Thoughts, in effect, are made up of internally reactivated traces of perceptions. Empiricists marshal philosophical arguments as well as linguistic, psychological, and neuroscientific evidence to support the perceptual vehicles thesis. Few, however, have distinguished among various strengths it might come in. We can distinguish these strengths as follows (using ‘percepts’ to mean ‘perceptual representations or copies thereof’):

1. Strong Global Empiricism (SGE): All thoughts are entirely composed of percepts
2. Weak Global Empiricism (WGE): All thoughts are partially composed of percepts
3. Strong Local Empiricism (SLE): Some thoughts are entirely composed of percepts
4. Weak Local Empiricism (WLE): Some thoughts are partially composed of percepts

Finally, opposed to all of these empiricisms is the concept rationalist thesis: *no* thoughts are even partially composed of percepts (Fodor, 1975); that is, thought is entirely amodal.²

Classical empiricists like Hume and Locke favored SGE, and their modern descendants such as Prinz and, at times, Barsalou join them in advocating this thesis. Others, e.g., Goldstone & Barsalou, only propose to defend a weaker view: ‘our position is that abstract conceptual knowledge is indeed central to human cognition, but that it depends on perceptual representations and processes, both in its development and in its active use. Completely modality-free concepts are *rarely, if ever*, used, even when representing abstract contents’ (p. 146, emphasis mine). This might indicate commitment to WGE or SLE. Some other theorists

¹ Spelled out more thoroughly, the thesis should include both perceptual and *motor* representations, but I will mostly discuss perception here. Also, combinations of copies of perceptual representations should be allowed to be concepts as well. The main argument of the paper is unaffected by these omissions.

² The weaker grades of empiricism distinguished in (2)-(4) can also be seen as weak forms of rationalism, since they allow that there is more to conceptualized thought than copies of perceptual representations. Moving away from empiricism involves moving closer to rationalism. To keep the terminology simple, though, I will consider these mainly as ways of stating different empiricist claims.

who are broadly inspired by empiricism are harder to classify (Glenberg & Kaschak, 2002; Glenberg & Robertson, 2000). In this paper, I'll focus on the neuroscientific evidence for what I take to be the boldest and most interesting empiricist claim, namely SGE; it is this that I refer to as 'the perceptual vehicles thesis' from here on.

Empiricists are committed to the idea that the neural correlates of perception coincide, at some level of analysis, with the neural correlates of thought. The first question I will address here, then, is whether neuroanatomical, imaging, and single-cell recording studies, as well as other arguments drawing on specifically neurobiological evidence, support SGE. (I set aside consideration of possible purely psychological evidence for SGE in this discussion.) I will argue that once we attend to the way that control of thought and behavior is causally structured in the brain, we can see that this strong claim is unsupported. A second question I will address is a more general one: what sorts of constraints should be imposed on the search for the neural structures that realize thoughts and other cognitive states? This is an important issue for empiricists and non-empiricists alike. Finally, I will consider whether the evidence offered for SGE might support a weaker, and more plausible, sort of mixed empiricist thesis. I will conclude by assessing the prospects for such weaker forms of empiricism.

2. Discovering vehicles

Thoughts and concepts, *qua* representational states, can be viewed under at least two aspects: what they represent (what their intentional content is), and what sort of structure does the representing (what the vehicle of that content is). No doubt talk of 'vehicles' and 'content' is metaphorical, but the basic distinction should be clear enough. That Damascus is in Syria, for instance, can be represented by a sentence or a map. The content is the same, the vehicle differs. Similarly in other cases: compare a photograph of the cat asleep on the bed with the sentence

‘The cat is asleep on the bed’, or compare a linguistic representation of the fact that world oil production is declining this decade with a graph of world oil production. Graphs, maps, photographs, sentences, and other representational systems provide different vehicles for representing content.³ Empiricism says that the vehicles of perception, whatever they are and however they are structured, are sufficient to carry all thinkable thought contents.⁴

The orthodoxy in philosophy of mind holds that mental states (properties, events, processes, etc.) are functionally individuated. Functionalism as a metaphysics of mind can be spelled out in various ways: the commonsense causal functionalism of Lewis and Armstrong, the machine functionalism of Putnam, the teleofunctionalism of Lycan, and so on (see Polger, 2004 for review). These different approaches have in common that they take mental states to be constituted by certain of their causal relations to other mental states, to bodily movements and stimulations, and to the wider environment. This is of a piece with the general reductionist nature of functionalism: mental states are just states that are appropriately situated in the causal network, and this network as a whole can ultimately be characterized just in terms of nonintentional properties and relations.

This functionalist metaphysics leads naturally to a two-stage methodology for discovering the nature of thoughts. In stage one, we spell out the causal role that defines the state type itself—that gives its individuation conditions. Whether this causal role is determined a priori (by conceptual analysis, say), a posteriori, or by some mixture of the two isn’t important.

³ I have been supposing that these different vehicles *can* represent the same content. Some, e.g., Haugeland (1991), have disagreed with this, saying that different representational genera (he surveys sentences, images, and holographic representations) do not in fact carry the same kind of content. Without going as far as Haugeland, we can agree that photographs, for instance, may encode more content than do sentences: a photo of the cat represents her fur as being a certain color, her head as being tilted at a particular angle, and so on. None of this is represented in the sentence. Still, both the photo and the sentence seem to represent the cat as being asleep on the bed, so there is at least some overlap in content. I think that none of the arguments I will present against empiricism here depend on any strong assumptions about the ability of different vehicles to encode the (partially) same content.

⁴ It does not, though, claim that all thought *contents* are just perceptual contents. In this respect contemporary empiricism differs from its historical antecedents. See the end of this section for further discussion.

In stage two, we look around to see what sorts of things stand in the causal relations specified in stage one; that is, we find the realizers, the things that play the role that individuates the state type.⁵ In the case of thoughts and concepts, we need to spell out the causal role that distinguishes conceptualized thought proper, then discover what sort of states realize that role.

Saying just what constitutes the functional ‘essence’ of concepts is hardly trivial.⁶ Focusing mainly on the criteria used by psychologists yields the claim that concepts are states that are centrally causally implicated in categorization and inference; that is, they are states that function in grouping together objects under a common heading, and in projecting from something’s being a member of one class to its having certain further properties. These are clearly fundamental psychological acts. There are other criteria as well: concepts need to be able to combine into larger structures, they need to be publicly shareable, they need to have the right sort of representational content, they need to be learnable, and so on (see Murphy, 2002 for further discussion of the explanatory function of concepts). But the role delimited by categorization and inference is the weightiest factor in the causal profile that defines concepts as a unified object of study. Of course, if the structures that are operative in categorization failed to satisfy other component of the causal role, we would have a problem. We might then need to look for some other states that *could* meet all of the components, or revise the description of the role itself so that it picked out some more narrow class. But here I will assume, to keep the discussion simple, that there is no problem about whether what explains categorization can also explain conceptual combination, intentional content, and so on.

⁵ Realizers of one role may themselves be functionally specified as well, of course (Lycan, 1981). Some have argued that all scientific theories, all the way down to basic physics, individuate their properties and kinds functionally—no science tells us about the true intrinsic natures of things (Shoemaker, 1981).

⁶ One might argue over whether we should begin with the analysis of concepts and then analyze thoughts, or proceed in the other direction. I will assume that the most fundamental fact about concepts in the present context is that they are constituents of thoughts. Given this relationship, it makes no difference whether we begin with the role of concepts or thoughts: either one can be defined in terms of the other.

Let us focus, then, on the empiricist argument for the perceptual vehicles thesis. Given the two-stage methodology, the case must be that the neural structures that are most causally significant for explaining the target processes (categorization and inference) elucidated in the concept-role are the very structures that support perceptual representations. An independent way of identifying perceptual representations is to say that they are representations used by sensory systems, defined as a creature's dedicated input systems (Prinz, 2002, p. 115). Given this specification, we can first isolate perceptual representations, then see that the same structures in which these representations are realized are also implicated in the concept role.

So stated, though, empiricism faces a puzzle. If thought vehicles are perceptual vehicles, how are thinking and perceiving distinguished? Why isn't any activation of the perceptual system *de facto* a deployment of concepts? A final condition is needed to draw this crucial distinction. To avoid the conclusion that any act of perceiving an object counts as an act of thinking proper, Prinz proposes that 'concepts are spontaneous: they are representations that can come under the *endogenous control* of the organism' (emphasis mine); 'concepts can be freely deployed in thinking, while certain perceptual representations are under environmental control' (Prinz, 2002, p. 197). Perceptual representations become concepts, then, when they are capable of being deployed under endogenous control.

The notion of endogenous control is, unfortunately, not a perfectly clear one. For every particular case in which a representation is activated by an internal cause, we could potentially trace that cause's history back to causes outside of the organism as well. Perhaps what is meant is something like this: representations are endogenously controllable when they can be caused by organism-internal states that aren't themselves reliably activated by any particular external stimuli. The endogenous control state itself must be independent of the stimuli presently

impinging on the creature. No doubt more could be done to clarify this notion, but since nothing will hang on it, we can set these objections to one side. The general point is that if a representation can *only* be tokened by a perceptual encounter with an object or stimulus, it is a mere percept; if the same representation can be tokened by organism-internal causes, it is a concept. So the criterion of conceptuality is modal: concepts are perceptual representations that *can* be deployed under endogenous control.⁷

3. Convergence zones and the locus of control

Empiricists maintain that the structures activated in categorization that are capable of being deployed endogenously are purely perceptual, i.e., those that are proprietary to some dedicated input system (*or* copies thereof; I ignore this disjunct for the moment, but it will become important later). I'll argue that there are cell assemblies that are plausibly implicated in categorization and that are endogenously deployable that are nevertheless not perceptual representations. If there are, then by the empiricist's reasoning, these must be counted as conceptual. So not all concepts are composed of purely perceptual vehicles, and SGE is false.

Interestingly, some of these assemblies are discussed by Prinz himself: they are so-called *convergence zones* posited by Damasio and colleagues in a number of papers (A. R. Damasio, 1989; A. R. Damasio & Damasio, 1994; H. Damasio, Tranel, Grabowski, Adolphs, & Damasio, 2004). Convergence zones are neural ensembles that receive projections from earlier cortical

⁷ Concept empiricism is not necessarily committed to the problematic empiricist doctrines mentioned at the beginning of this paper. For example, it is not verificationist. Percepts represent perceptual properties, but they may also be used to represent categories that transcend perception. One way to achieve this is by adopting an informational theory of content on which structures represent what they are nomically connected with (for elaboration, see Prinz, 2006). Briefly, on such a theory distal properties whose instances cannot be directly perceived can be represented using perceptual vehicles that correspond with their proximal traces. Whether such a theory of content is ultimately successful is not something I will deal with here; I mention it only to block the immediate charge of verificationism. Neither is concept empiricism committed to associationism. As a claim about representational vehicles, it potentially allows that many kinds of processes may operate in cognition to transform, select, store, and combine representations. These need not be limited to Humean or connectionist associations. The perceptual vehicles thesis is intended to be neutral with respect to how concepts get their intentional/representational properties, and with respect to the range of processes that operate on those vehicles.

regions (e.g., lower sensory areas), contain feedback projections onto those earlier layers, and feed activity forward into the next highest layers of processing. These zones can also refer back to zones earlier in the processing stream, so that higher-order zones may be reciprocally connected with multiple lower-order zones.⁸ The functional role of convergence zones is to orchestrate the reactivation of lower-order activity patterns. A particular zone can become sensitive to the co-occurrence of a particular pattern of activity in the region that feeds it in such a way that that activity pattern can later be reinstated endogenously by activity in the zone itself.

The fact that convergence zones can engineer the re-deployment of perceptual representations of concrete entities and events gives the empiricist some support, since it shows that percepts can be activated off-line during bouts of cognition (e.g., drawing inferences about objects, naming them, recalling their properties, and so on). However, this falls short of showing that only percepts serve as concepts. The problem is that convergence zones *themselves* must be activated in order to orchestrate perceptual re-enactments. Zones may receive downward connections from a variety of areas, including prefrontal cortex, cingulate, basal ganglia, thalamus, and so on. So neuronal ensembles in convergence zones satisfy the endogenous deployability condition.

Do they satisfy the categorization condition, however? It is hard to see why they don't. In any particular episode of categorization, an enormous range of neural structures is implicated. Everything from low-level sensory receptors at the retina and skin through thalamic relays and associated subcortical structures, sensory cortices, and various prefrontal regions receive activation; and this list doesn't include the output side involved in generating behavioral responses. Included in this vast pattern of activation are convergence zones: if a person is trying

⁸ Terminological note: I follow neuroscientists' usage in calling a cell assembly higher-order when it receives projections from assemblies that are closer to the sensory periphery, and lower-order when it projects forward to regions that are further from the sensory periphery. That is all that these terms mean in the present context.

to recognize a perceived object, zones are activated that aid in retrieving its name and associated properties. If an object is being perceived for the first time, zones are receiving inputs that train them on the particular pattern of co-occurring perceptual properties. If no object is presented, but an inference must be made about a class of objects, zones are again re-activated as part of the process of knowledge access. So in many paradigmatic categorization tasks, convergence zones are implicated, insofar as they are central to retrieving the representations of perceptual features of objects.

Convergence zones, in fact, seem to make a unique contribution towards representing conceptual contents. Individual representations of perceptual features are realized in modality-specific sensory cortices. When some set of those features are active, we can say that the system is representing certain properties of a perceived or remembered object. However, complete concepts and thoughts are composed of more than just co-activated features. To see this, consider the set of features RED, ROUND, BLUE, SQUARE. This set is indeterminate between representing a red round object and a blue square object and representing a blue round object and a red square one. The mere co-occurrence of those features doesn't decide which concepts are bound together as representing the same objects; this is the familiar 'binding problem' as it is discussed in the neuroscience literature.

Convergence zones are responsible for re-activating perceptual features, but moreover it has also been claimed that they represent 'the combinatorial arrangements (binding codes) which describe their pertinent linkages in entities and events (their spatial and temporal coincidences)' (A. R. Damasio, 1989, p. 39). So it is the activity in the relevant convergence zone that determines which of the two possible pairs of bindings of adjectival and nominal concepts is the correct one in the example given above. This semantic function is essential, since it makes the

difference between a mere loose bag of concepts that are co-tokened and an articulated representation with predicative and propositional structure. On this interpretation, convergence zones are essential to representing the most abstract, amodal aspects of thought contents, namely their logical skeleton.⁹

Given that they satisfy the causal-role conditions on being concepts as well as perceptual representations do, convergence zones have an equal right to be seen as (part of) vehicles of thought. This possibility threatens Strong Global Empiricism. If all thoughts have some amodal component, the strongest claim that can be established is Weak Global Empiricism. In discussing this objection, Prinz says:

Convergence zones may qualify as amodal, but they contain sensory records, and they are not the actual vehicles of thought. Convergence zones merely serve to initiate and orchestrate cognitive activity in modality-specific areas. In opposition to the rationalist [amodal representation] view, the convergence-zone theory assumes that thought is couched in modality-specific codes. (Prinz, 2002, p. 137)

First, this seems to overstate the commitments of the theory as Damasio et al. express it. It isn't a necessary part of believing in convergence zones that one believe the perceptual vehicle thesis. A slimmer commitment will do, namely the thesis that perceptual representations are causally necessary for carrying out certain kinds of categorization and inference involving some categories. It might be useful to re-activate perceptual records of one's encounters with dogs in deciding what shape a German shepherd's ears are, or records of frog encounters in deciding whether they have lips. This shows that re-enacted perception is useful, and perhaps necessary,

⁹ I ought to stress that this interpretation is highly conjectural. I'm not endorsing it wholeheartedly here, just raising it to illustrate a possible role that convergence zones might fill.

in many contexts. Convergence zones may be instrumental in making this re-enaction happen. But it's a further step to thinking that the vehicles of thought are *just* the re-enacted perceptions.

Second, in the present context it is question-begging to assert that convergence zones aren't vehicles of thought. As we've seen, the higher level non-perceptual states that cause the re-enactions have an equal right to be seen as vehicles along with the perceptual states themselves. This might seem to serve empiricists a qualified victory: *some* thoughts are made of percepts. But in the very same episodes of occurrent thought, there are non-perceptual representations active as well. The most radical concept rationalists might be unhappy with this conclusion. Even so, the strongest version of the perceptual vehicles thesis seems similarly unsupported.

The empiricist has another possible reply at this point. He might maintain that convergence zones aren't representational at all. Rather, they are *mechanisms* for generating representations. This is hinted at in the claim that they 'orchestrate' re-activation of perceptual states. Barsalou, too, suggests that '[a]lthough mechanisms outside sensory-motor systems enter into conceptual knowledge, perceptual symbols always remain grounded in these systems. Complete transductions never occur whereby amodal representations that lie in associative areas totally replace modal representations' (1999, p. 583). While this claim is cautiously hedged, it suggests a reply similar to Prinz's: *mechanisms* outside of the senses enter into concepts.¹⁰ Empiricists might, then, propose a strong distinction between (i) representational states and (ii) non-representational mechanisms that token and transform those states. In its most general form, the *mechanism strategy* (as I will call it) involves arguing that any putatively non-perceptual representational activity can be re-interpreted as the activity of a mechanism.

¹⁰ The quoted passage is, in fact, somewhat more hedged than this, since it seems to admit the possibility that there *are* amodal symbols somewhere in the mind/brain. Prinz at times seems to want to avoid even this conclusion. For more on this issue, see the discussion of multimodal cells in the next section.

This strategy faces several problems, however. First, it isn't clear that the strong distinction mentioned above can be sustained, because some mechanisms might *also* be representations. Consider a mechanism that implements a transition from symbols shaped like 'P→Q' and 'P' to symbols shaped like 'Q'; that is, a mechanism that implements a transition corresponding to the application of *modus ponens* (MP). One might argue that a system containing this mechanism, in virtue of containing it, *implicitly represents* the rule of MP. Cummins (1989) argues that computational systems frequently include such implicit representations. Generally, systems may implicitly cognize a rule by having a mechanism the operation of which can be semantically interpreted as corresponding to the application of the rule. In virtue of having a mechanism by which complex symbols can be transformed in the right sort of way, the system grasps the rule of MP. Implicit representation shows one way in which mechanisms might also be themselves representational. Since some mechanisms (e.g., the one that transforms symbols in accord with MP) implicitly represent, calling convergence zones mechanisms doesn't automatically show that they aren't also representations.

Further, cognitive and neural mechanisms are complex structures that may themselves contain various sorts of *explicit* representations (for analysis of the notion of a mechanism, see Machamer, Darden, & Craver, 2000). Mechanisms contain parts that operate together in a spatiotemporally organized way to subserve a certain function. These parts themselves may be explicitly representational. For example, there might be an amodal representation DOG tokened in a higher-order convergence zone that functions to re-activate perceptual traces of dogs in lower order sensory regions. These amodal representations might be part of mechanisms for the directed retrieval of perceptual information in the service of carrying out particular cognitive tasks that are best addressed by using this information. So, as this simple example illustrates, the

mere claim that convergence zones are mechanisms doesn't by itself establish their non-representational credentials (I will develop this objection further in Section 6).

Moreover, it isn't clear that the mechanism strategy is consistent with some empiricists' preferred informational theory of content (on informational semantics, see Dretske, 1981; Fodor, 1990).¹¹ On informational psychosemantics, roughly speaking, a state's content is the condition that it nomically covaries with. In the case of percepts, this is the state that nomically causes them; in the case of motor representations, it is the state that they nomically cause. Convergence zones, as understood on the mechanism strategy, are analogous to motor representations: they function to reliably bring about a certain (neural, not behavioral) effect state, namely the re-enactment of lower-order perceptual representations. If this relationship is nomic, it seems that on an informational account it would be difficult to resist the conclusion that they represent. Just what their content is isn't entirely clear; perhaps they represent the perceptual representations that they function to produce, or perhaps they represent the properties that those percepts represent.¹² In any event, whatever one says about the precise nature of their content is less important than the fact that they can plausibly be seen as non-perceptual representations.

One might try to block the conclusion that convergence zones are amodal representations by appealing to a more stringent notion of representation. It might be that nomic covariation is insufficient for representation. Perhaps having the right sort of function is also required. Representations might need to play the role of 'standing in' for what they represent when those

¹¹ Prinz is most explicit in adopting informational semantics. But it is worth noting that informational approaches are consistent with what many neuroscientists seem to believe about neural representation. The standard empirical method for assigning content to a cell assembly is to locate the stimulus conditions that cause it to fire above its normal baseline. These conditions can be regarded as the properties that the cell is 'locked onto', in the jargon of informational semantics. Receptive and projective field mapping are, then, ways of empirically discovering neural content on the assumption that this content is fundamentally informational.

¹² Note that not everything that represents what a perceptual representation does needs to itself be a perceptual representation. So even if convergence zones could represent perceptual properties, they wouldn't therefore be perceptual states. Perceptual states are, for the empiricist, defined as being part of perceptual systems. But there is no apparent reason why the content of those states can't be duplicated elsewhere in the brain.

things themselves are absent. In that case, if convergence zones always co-occur with lower-level perceptual states, they cannot *represent* those states (or those states' content). This criterion of representation, though, may be too stringent. It runs the risk of ruling out motor representations, for example, which by design play an active role in producing the movements that they represent. The challenge in making good on this anti-representationalist maneuver, then, is to find a condition on representation that rules in motor representations but rules out convergence zones. This is, at least, a significant challenge.

For a variety of reasons, then, the mechanism strategy seems unpromising. There is a strong case for seeing convergence zones as non-perceptual representations that play a central role in categorization processes.

4. Modality-responsiveness and perceptual representation

At this point we can see that more than purely perceptual representations seem to be implicated in the overall categorization process. But empiricists cite other evidence to support the perceptual vehicles thesis. They might argue, for example, that there are no genuinely amodal cells anywhere in the brain: all neurons are plausibly associated either with perceptual processing or with interfaces among perceptual systems. While amodal cells might be taken to be the hallmark of non-perceptual representation and processing, empiricists can argue that any evidence for purportedly amodal cells could be reinterpreted as being evidence for bimodal or multimodal cells instead. Since bimodal or multimodal cells can be seen as being perceptual representations, their existence doesn't threaten empiricism.

Discussion of modality-specific neurons arises in the context of a potential objection to empiricism. Newborn children are able to map stimuli perceived in one sense modality onto stimuli perceived in another. So, for instance, at one month of age they can map objects felt with

their tongues (such as distinctively shaped pacifiers) onto seen objects (pictures of those pacifiers) (Meltzoff & Borton, 1979). One candidate explanation of this ability is that there are cells in regions such as the superior colliculus that function to map stimuli from one sense modality to another. An anti-empiricist might contend that these are genuinely amodal cells that plainly function in categorizing stimuli (deciding that *what is felt* is the same as *what is seen*, for instance). If there are amodal cells that have this function, then one might conclude that there are genuine amodal neural representations at work in cognition.

It is crucial to guard against a conflation between *a cell's being modality-responsive* and *its being a component of a perceptual representation*. Being modality-responsive is a matter of whether a stimulus of a certain type causes a neuron to activate, or causes it to activate preferentially. Prinz characterizes the senses as 'dedicated input systems', meaning that they (1) function to receive inputs from outside the brain, (2) are distinguished on the basis of the different physical magnitudes that they respond to, and (3) are housed in separate neural pathways, as identified by neuroanatomical and functional considerations (Prinz, 2002, pp. 115-7). So while modality-responsiveness is necessary for being a perceptual representation, it isn't sufficient. Even if it turns out that there are widespread unimodal cells in the brain, more is needed to show that these are components of perceptual representations. Specifically, one needs to show that these cells are part of separate, internally distinguished neural systems that have the function of processing physical input magnitudes.

Prinz's response to the challenge from amodal cells seems to overlook this caveat. though. He suggests two possibilities: the empiricist could either argue that apparently amodal cells are really modality specific, or argue that they are multimodal. But showing that these cells are genuinely specific in their response pattern to stimuli from a certain class of magnitudes

doesn't show them to be perceptual representations unless it's also established that they are part of a distinctive neural system that has the right sort of input function. Distinguished response patterns are only one component of perceptual representation.

The multimodal cell response needs further consideration. Here the proposal is that multimodal cells have a foot in more than one sensory system and function to map information from one system onto another. In the case of intermodal transfer it might be reasonable to think that a relatively simple interface might do to link distinct sensory systems. But it is questionable whether every body of multimodal cells is implementing such a (relatively) simple mapping. So even if some multimodal cells can be treated as boundaries or interfaces between sensory systems, not all of them plausibly can.

These larger populations of multimodal cells may comprise distinct, *non-perceptual* processing systems. Given the size of multimodal cell populations, in fact, it seems quite likely that there are non-perceptual processing systems in the brain. Large portions of the prefrontal cortex are dedicated to multimodal processing, including dorsolateral, superior, and inferior regions, pars triangularis, and pars orbitalis (Kaufer & Lewis, 1999). It is implausible that these are simply serving as interfaces between sensory systems. The greater the size and complexity of the multimodal region that mediates between sensory systems, the less it appears to be a simple interface or boundary and the more it appears to be a distinct, non-sensory system dedicated to carrying out a separate cognitive task. So the multimodal cell response does not generalize: not every multimodal cell is just part of an interface.

This last claim can be strengthened by considering recent evidence that multimodal cells are widely distributed in the brain, even in so-called primary sensory regions (this evidence is reviewed extensively in Ghazanfar & Schroeder, 2006). These multimodal cells are frequently

specialized to fire in the presence of congruent multisensory stimuli. For instance, certain cells in the primary auditory cortex of monkeys fire preferentially when a face and a voice are present together, and the so-called fusiform face area (located in a traditional ‘visual’ cortical region) has been found to activate to familiar voices as well as familiar faces.

These findings also challenge the notion that ‘each type of [perceptual] symbol becomes established in its respective brain area. Visual symbols become established in visual areas, auditory symbols in auditory areas, proprioceptive symbols in motor areas, and so forth’ (Barsalou, 1999). One prominent conception of what makes something a perceptual system (as noted, a conception endorsed by Prinz) is its responsiveness to a unique physical magnitude. But not every representational tokening in a visual region may count as a ‘visual symbol’ if some visual cells have their response properties modulated by non-visual stimuli as well. None of this is to say that an empiricist might not be able to construct a definition of a sensory system, and hence of perceptual representations, that is compatible with widespread distribution of cells that respond to different classes of physical magnitudes. But if traditional unimodal sensory regions in fact contain a large percentage of cells that respond to bimodal or multimodal stimulation—hence, that respond to many different kinds of physical magnitudes—the notion of a pure or dedicated sensory system on which empiricism depends may have to be revised.

To summarize, then, there are three separate roles that multimodal cell populations might be playing: (1) they might be interfaces between sensory systems; (2) they might be non-sensory processing systems; or (3) they might be parts of sensory systems that are directly modulated by activity in other sensory systems. Insofar as roles (2) and (3) predominate, multimodal cells cannot easily be interpreted in a way that is compatible with empiricism.

5. Copies and the collapse of empiricism

Earlier I noted that the perceptual vehicles thesis is disjunctive: concepts are either percepts or *copies* thereof. The notion of a copy now needs closer scrutiny. I've argued that at several crucial functional regions in the brain there are cell assemblies active in concept-central tasks that aren't employing perceptual representations. The grounds for thinking that these representations are non-perceptual is that they are not the proprietary currency of any of the senses, where these are understood as dedicated input systems in the brain. Empiricists might still be able to accept this and argue instead that cells in convergence zones or regions that are identified with working memory such as prefrontal cortex can be regarded as manipulating copies of perceptual representations. The notion of a copy, when scrutinized, turns out to undermine the substance of the contemporary empiricist thesis.

Copying can be construed in several ways. In classical empiricism, copies of impressions—ideas—necessarily preserved their phenomenological qualities modulo a lesser degree of 'vividness'. Crudely, a copy was a state that appeared in consciousness as a 'washed-out' perception. In contemporary empiricism, though, phenomenological properties aren't necessarily preserved in copies; indeed, phenomenology is not even necessary for something's being a perceptual representation, since a lot of perceptual processing is non-conscious. Copying is, broadly, a causal process that produces new representations from old ones. This suggests that copies obey an etiological condition: a copy has its causal source in the representation that is its original. A chance resemblance in two structures isn't enough for one to be a copy of the other.

In addition to etiology, there is a common content condition: For one representation to be a copy of another, it is necessary that it share content with its original. If copying doesn't preserve content, it isn't distinguished from inference and other processes of information transformation. But shared content isn't sufficient for something to be a copy. This is because the

same content can be re-encoded into different formats—different kinds of vehicles. Consider an analogy. A low-fidelity hand-drawn sketch made from an ornately decorated treasure map might be a copy of that map even if it omits much of the detail in the original map. It can be a copy even if it is a different size, and if it replaces, say, a picture of a treasure chest with an ‘X’ to mark the location of the loot. But a detailed verbal description of the location of the buried treasure is not a copy of it. The description might preserve the map’s content, but that isn’t enough for it to count as a copy. Some degree of vehicular similarity is needed as well. In particular, we need at least to have the same vehicle *type* in both the original and the copy. Both original and copy need to be maps, and maps that contain the same content. Small details in the way that content is marked on each map may not count as differences in either content or vehicle (unless they have some significance in the representational scheme, or in the way the vehicle is processed, in which case they are clearly relevant to determining the similarity of an original and its putative copies).

Prinz seems to endorse this conclusion. He suggests that a necessary condition for something’s being a copy is that it preserve such properties of the vehicle’s format: ‘Imagine, for example, that a visual percept is a pattern of neural activity in a topographic map corresponding to the visual field. A stored copy of that percept might be a *similar pattern in a topographic map stored elsewhere in the brain*’ (Prinz, 2002, p. 108; emphasis mine). Let’s assume, then, that it is necessary and sufficient for something’s being a copy of a percept that:

1. it has its causal source in the original percept;
2. it has the same content as the original; and
3. it is similar enough in its vehicular properties to the original perception.

The question for the empiricist now becomes how widespread these copies are in the brain.

At least some brain areas that process sensory information do so in a different way than do primary sensory cortices. The superior colliculus, for instance, receives visual, tactile, and auditory sensory inputs and is implicated in directing eye movements towards targets that can be localized using all of these sensory cues. It might seem that the colliculus contains perceptual copies, since it is roughly organized as a series of maps of different sensory spaces laid on top of one another. In the case of vision this map is similar to others found in primary visual cortex. But on closer inspection things aren't so simple. The auditory map in the colliculus, for example, seems to represent the origin of sound sources in space around the organism. Primary auditory cortex, though, *doesn't* contain a spatial map of sound sources. Rather, it encodes a tonotopic map in which sound frequencies are mapped onto groups of neurons (low to high frequencies are mapped caudally to rostrally; see Middlebrooks, 2000). The primary auditory map represents frequencies, not locations of sounds. Since the auditory map in the superior colliculus is not only *structured* differently from the map in primary auditory cortex but also represents different *content* than that map, it cannot be a *copy* of an auditory representation but must be seen as a distinct structure for using auditory information in a task-specific way (probably integrating multisensory information for guidance of eye movements). It would not be entirely surprising if later representations of perceptual information employed specialized formats more suitable to the particular task they are involved in carrying out.

Another region that might be thought to contain copies of perceptual representations is prefrontal cortex (PFC), particularly the dorsolateral or anterior pole of PFC. It is not known exactly how to characterize the activity of PFC, but it has been implicated in retrieval from episodic memory, intentional direction of attention, planning, and setting and maintaining goals beyond the immediate context (see Ramnani & Owen, 2004 for review). PFC also receives

reciprocal connections from many sensory cortices. Importantly, PFC neurons seem to represent perceptual information differently than do the sensory areas that feed them. For example, Freedman, Riesenhuber, Poggio, & Miller (2001) trained monkeys to discriminate cats from dogs based on computer-generated images. A range of 'cat' and 'dog' images was created by morphing the focal cat and dog images together in varying percentages. Recording from cells in the monkeys' ventrolateral PFC revealed that some cells fired selectively for all dog stimuli, some for all cat stimuli. Moreover, when a monkey trained on this task was retrained on three new perceptual categories, these cells changed their response properties to fire selectively for the members of the newly relevant category.

This suggests that cells in some PFC regions may be adaptive, changing their representational properties with the task demands (Duncan, 2001). While there may be distributed representations of the perceptual properties of cats and dogs in various sensory cortices, it is less clear that there are specific cell populations that co-vary with particular categories of dogs and cats as such. These would be akin to the fabled 'grandmother' cells that encode memories of particular individuals (e.g., Grandma) and categories. Perhaps one role of PFC is to assemble these transient 'grandmother'-like structures on-line. Regardless of whether this speculation is correct, the point of interest is just that PFC may contain neural vehicles that differ from those deployed in sensory systems. Hence it cannot just contain copies of perceptual vehicles, even if it contains re-representations of categories that are also encoded perceptually.

Finally, evidence that neurons in PFC represent information differently than do perceptual systems comes from Rao, Rainer, & Miller (1997). It is widely known that in visual processing, information about an object's identity is processed separately from information about its spatial location. An alternative way to characterize this split is as being between information

that is used for the spatial control of behavior with respect to an object and information used to categorize an object (Milner & Goodale, 1995). However it is characterized, the former sort of information is processed in the dorsal stream of the visual system and the latter in the ventral stream.¹³ Separate representations and mechanisms operate on each kind of information.

However, populations of neurons in the dorsolateral PFC are responsive to both information about an object's spatial location *and* information about its category membership. These cells seem to integrate information from both visual processing streams. But this means that they cannot simply be *copies* of earlier perceptual representations, since the visual system represents this information separately. Neither are they easily seen as simple interfaces between two visual streams, since these cells are located in a highly multimodal region that engages in elaborate processing and retrieval of information for top-down control of behavior.

Ultimately, the copy proposal poses a dilemma for empiricists. Either copies are required to have a high degree of vehicular fidelity or they aren't.¹⁴ If they aren't, then, potentially any content-identical state can be a copy of a percept. But this undermines the central empiricist idea that thinking involves re-activating and re-using the very representations that are employed in perceiving objects and events. One might as well believe in an amodal language of thought on this conception of a copy. Further, it is far from clear what relevance the neuroscientific and psychological data about the use of perceptual representations could have, if copies need not themselves have similar structure or function as those percepts. Presumably the support that these studies give to empiricism is in showing that properly perceptual representations are

¹³ There may also be several separate types of information processed within the dorsal and ventral streams. These distinctions are coarse, but sufficient for the present discussion.

¹⁴ Of course, the notion of 'fidelity' here is a graded one, since it relies on the notion of *similarity* in neural vehicles. The issue is complicated by the fact that vehicles can be similar in some respects but different in others. A spatial map of visual space in one region of the brain may not behave the same as a map of the same space in another region of the brain, owing for example to differences in the local connectivity patterns within each region. When to count any two maps as being similar enough to one another to support the same representations is a difficult issue to decide.

recruited in a wide array of conceptual tasks. Moreover, if the vehicular similarity clause is dropped, empiricists would then be vulnerable to a charge they sometimes make against amodal theorists, namely that their theory can account for potentially any observations, and hence becomes drained of distinctive testable consequences. Perceptual representations can be independently identified, given a definition of what counts as a perceptual system. Copies of perceptual representations that differ in their vehicular properties, though, cannot.

On the other hand, if copies are required to be highly similar in vehicular properties to perceptual representations, then it's an open question how widely they are distributed throughout the brain. Many regions outside of the bounds of perceptual systems manipulate and process perceptual *information*. But, as the examples presented here suggest, they may do so using differently structured representational vehicles. So the distribution of copies may not be as extensive as empiricists require to establish the perceptual vehicles thesis. To establish the full version of the copy proposal, empiricists need to provide convincing evidence that the vehicular properties of neural populations in perceptual regions are duplicated elsewhere in the brain. To date, this evidence is lacking.

6. Refining neural correlates

I have raised a number of arguments against the perceptual vehicles thesis. The common structure of the arguments that I have been offering is, essentially, the following: in explaining the central functions of the conceptual system, we will need to appeal to many neural activation patterns beyond those in the perceptual systems; these neural activation patterns are not themselves perceptual representations or copies thereof; hence SGE cannot be correct, since it presumes that all of the representations involved in these central functions are perceptual.

Note that the form of argument given here doesn't show that we won't also, at least sometimes, need to explain people's reasoning and behavior by appeal to perceptual representations. (Perhaps we'll always need to appeal to the deployment of some perceptual representations or other; I am agnostic on this issue.) An empiricist, then, might reply to these arguments by saying that neural activity in non-perceptual systems shouldn't be counted as activity that partially realizes thoughts. That is, an empiricist might try to *carve off* part of the ongoing pattern of activity in the brain—namely, the part involving only perceptual representations—and argue that, while other brain regions might be causally involved in deploying these percepts, they aren't constitutive of thoughts themselves. On this strategy, what is needed is some way to draw a distinction between neural states that are (merely) *causally involved* in the production of a thought, and states that *constitute* a thought.

In principle this distinction might seem easy enough to draw. Isn't it obvious, after all, that there is a clear boundary around the occurrence of a particular thought here and now and everything else that is going on in my mind, body, and environment right now, including that thought's causes and effects? Mustn't there be such a boundary if there is a fact of the matter about what particular thoughts I am entertaining, both at a time and diachronically? Perhaps some intuitions say so. In practice, however, locating the boundaries around these thoughts as they are realized in the brain is considerably harder.

This task bears a powerful resemblance to the task of locating the so-called neural correlate(s) of consciousness (NCC). The logical structure of this task has recently attracted a fair amount of attention (Block, 2005; Chalmers, 2000; Noë & Thompson, 2004). Looking at the structure of this related debate can help to shed light on our present topic, which we might call the search for the *neural correlate of conceptual thought* (NCCT).

The job of finding an NCC, for some particular conscious phenomenal state type, involves locating a neural state the activation of which co-varies appropriately with the occurrence of the phenomenal type. Chalmers (2000) gives a general definition of an NCC as follows: ‘An NCC is a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient, under conditions C, for the corresponding state of consciousness’ (p. 31). So, for instance, if we are talking about the state of consciously experiencing a red patch in a region of visual space, consciously perceiving a horizontal line, or some other visual state that has a phenomenological component, the neural correlate of that state is nomically sufficient for the occurrence of that experience under the appropriate conditions.

Saying what makes a set of conditions appropriate is difficult. Chalmers notes that normal conditions are plausibly taken to require an intact (unlesioned, developmentally ordinary) brain, although it may undergo unusual stimuli (either external or internal, e.g., presentation of unusual stimuli via normal perceptual channels, microelectrode stimulation, or transcranial magnetic coil stimulation) during experimental procedures. Since brain lesions can induce widespread changes in the architecture and function of normal brains, he cautions against taking the conditions C to include them.¹⁵

Moreover, and significantly given our present concerns, a neural correlate must be a *minimal* state, that is, the smallest region the activation of which is sufficient to produce the phenomenal state in question. This is to rule out taking the entire brain of a creature to be some

¹⁵ For this reason, as well as considerations of space, I have omitted discussion of the role that evidence from category-specific deficits might play in establishing empiricism. Generally, the findings involve showing that lesions to perceptual regions produces selective deficits in identifying and reasoning about entire classes of objects (e.g., living things or artifacts). However, it is far from clear whether there is any consistent correlation between lesioning a region and production of a certain deficit; in addition, there is little standardization of the tests that demonstrate these deficits, so different researchers characterize them in different ways. For a review of the literature and discussion of models that might explain category-specific deficits, see Caramazza & Mahon (2003) and Humphreys & Forde (2001).

state's neural correlate. Clearly some activation pattern in the whole brain would suffice for the occurrence of a particular conscious state, but that would not be very informative about what features of the brain in particular were responsible for producing consciousness.

A final condition on this minimal neural state is that according to Chalmers it must be a *content match* for the conscious state of which it is a correlate. The neurons that are activated must represent the same content that is revealed in consciousness. For instance, if a neuron fires preferentially to a horizontal line in its receptive field, then it may provisionally be tagged as a representation of a horizontal line in such-and-such a location, and can potentially count as a correlate for the conscious experience of seeing such a line.

An empiricist might try to exploit the twin conditions of minimality and content matching in order to winnow down the NCCT. Here is one way this argument might go:

1. Content matching requires that the NCCT, no less than the NCC, carries the same content as does the thought of which it is a correlate.
2. While there is widespread neural activity during conceptualized thought, *only* the activity in perceptual regions carries the same intentional/representational content as the thought being realized.
3. By minimality, the NCCT should be identified with the smallest neural region that satisfies content matching.
4. Hence, only the activity in perceptual regions counts as the NCCT.

The argument is a valid one. Premises (1) and (3) are simply statements of the conditions on neural correlates adapted from Chalmers' discussion of NCCs, which aren't controversial in the present context. The key empirical premise in the argument is (2). We should consider, then, what arguments can be given to support it.

First, note that the definition of a neural correlate requires that the activity be sufficient for occurrence of a thought.¹⁶ But if we are considering just the intrinsic activity of a perceptual system at a time, this just raises the very problem for empiricism that was mentioned in Section 2, namely the problem of distinguishing thought from perception. There I followed empiricists in distinguishing concepts from percepts on the basis of whether they could be brought under endogenous control. As noted, though, this is a modal criterion. It can't tell us whether a particular pattern of activity in visual cortex is a *perception* of a dog or an internally re-activated perceptual state being deployed as a *thought* about dogs. The pattern itself doesn't determine which of these possible cognitive states is occurring. If that activity *could* have been generated endogenously, then those representations count as both conceptual and perceptual. But this won't tell us whether the organism is *now* using them to think or to perceive. Putting the point somewhat differently, the endogenous control principle can distinguish concepts from percepts, but not thinking from perceiving.

We might try extending the endogenous control principle here to distinguish occurrent thoughts from occurrent perceptions. *Actual* causal etiology might distinguish thinking from perceiving. Perhaps occurrent thoughts are those perceptual states that are actually caused endogenously, and perceptions are states that are caused by external stimuli. This won't quite work as a sufficient condition, though, since hallucinations are internally caused *perceptual* states. There is no dagger before me when I hallucinate one; the dagger appearance is caused by some state of my disordered brain. Moreover, it's not clear that this proposal works as a

¹⁶ In their critique of the notion of an NCC, Noë and Thompson (2004) argue that insisting on sufficiency presupposes an *internalist* notion of content, since on externalist notions of content, including informational semantics, factors outside of the neural state itself help to fix what it represents. So the mere occurrence of that state itself, in abstraction from its surroundings, need not constitute the tokening of a contentful representation. An analogous critique might be raised in the present case, but the argument that I am presently making applies even if Noë and Thompson's problem could be solved.

necessary condition, either, since not all thoughts must actually be endogenously caused. Perceptual beliefs are occasioned by perceptions, which in turn are actually caused by the environment, but there is nevertheless a difference between occurrent perceptual belief and perceiving. So actual etiology seems unable to draw this distinction.

This problem about distinguishing thinking from perceiving is relevant to empiricists who want to exploit minimality in order to locate the NCCT in perceptual regions, since activity in those regions might be sufficient for *either* perceiving or thinking. But perhaps this disjunction doesn't matter to the issue of whether those regions are NCCTs, as long as the activity in these perceptual regions is a content match for the thought that the creature is entertaining. The fact that we don't have a clear criterion for distinguishing perceptual system activation that is thinking from that which is perceiving might be a non-issue as long as the underlying neural state in each case is the *only available* content match for the thoughts that we are trying to correlate with the underlying brain states. Rather than pursue the problem of distinguishing thought and perception further, then, I will turn to the issue of the uniqueness of content matching.

Unfortunately for the empiricist, we have already seen, in the earlier discussion of convergence zones (see Section 3), that content-bearing states are potentially widely distributed throughout the brain. This means that there are many alternative candidates for the NCCT. Higher-order convergence zones provide one instance, regions in prefrontal cortex provide another (these may not be exclusive categories). Both of these regions potentially carry representational content; indeed, as the Freedman et al. study showed, cell assemblies in prefrontal cortex can adapt their response properties to different categories depending on the task demands.¹⁷ Suppose, then, that we have simultaneous (or nearly so) activity in both a higher-

¹⁷ Although I should note that this depends on convergence zones being representations of the appropriate intentional content. If they either are not representations or are representations only of the lower-order neural states

order region such as a prefrontal cell assembly and a lower-order perceptual region, say some part of visual cortex. Perhaps the lower-order activity is under the active causal control of the higher-order state, thus satisfying the endogenous deployment condition. Are there *any* grounds for decisively singling out the perceptual activation as the minimal bearer of intentional content in this case as opposed to the higher-order state that is causally directing it?

I suggest that we don't have such grounds. It seems equally possible that the intentional content of thought is represented in the activated higher-order regions as in the lower-order perceptual regions. To see how this might be possible, consider one model I sketched earlier in discussing what I called the mechanism strategy. On this model of neural processing, there is an amodal representation of a certain category, say DOG, in some extra-perceptual area. This representation is deployed as part of a process of making some inferences about dogs, e.g., deciding whether dogs have spleens. The functional role of this representation might be to guide further mental processing in deciding this question by coming to judge either DOG HAVE SPLEENS or DOGS DON'T HAVE SPLEENS. Suppose now that as part of the mental processing that goes into answering this question one activates some mental images of dogs—perhaps culled from a television program on veterinary medicine or a textbook of dog anatomy, for example. These images might be realized neurally as patterns of activity in some part of visual cortex, and these patterns might represent dogs, their spleens, and so on. By manipulating these images—say, comparing the visual representation from the veterinary program with a stored visual

themselves, then this line of argument fails. But as I argued earlier, we *have* reason to think that they are representations. Reason to think that they represent what lower-level perceptual states represent might come from other cases in which representational content is 'borrowed'. For instance, a child may learn an elephant concept by encountering not elephants themselves but words and pictures of elephants. The concept learned borrows its content from these representations. Perhaps the same thing happens with convergence zones: they represent what the perceptual representations do, rather than the vehicles themselves. If this is true, then they are candidates for the NCCT.

representation of a human spleen—one might be able to make the appropriate inference. If there is an appropriate visual match, then one comes to judge that dogs do in fact have spleens.

I'd like to stress that I'm not offering this as a detailed and realistic model of how categorization judgments are processed in the brain, although something like it might not be too far from the truth. Often we seem to use visual images to solve problems, and using such images involves re-activating parts of perceptual areas (Kosslyn, 1994). Some evidence for a structuring of control and retrieval systems in the brain akin to the one proposed here comes from Rowe et al., who found that activity in dorsolateral PFC (Brodmann area 46) is associated with retrieval of items from memory to guide behavioral responses (Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000). For a perspective on the role of PFC similar to the one sketched here, see Miller, Freedman, & Wallis (2002). If this simple model is correct, we have *two* candidate regions that might serve as the NCCT: the perceptual regions themselves, and the higher non-perceptual cortical regions that are orchestrating the activity in perception. On this interpretation of what is happening in this categorization task, the intentional contents DOG, SPLEEN, etc., are represented in both modality-specific and amodal neural regions. And if this is the case, we cannot privilege *either* region as the unique neural vehicle of thought, given the constraints of minimality and content matching.¹⁸

Generally, the empiricist's strategy of locating content-matching neural states solely in perceptual regions faces problems because categorization and inference are typically processes that involve activating goal representations, searching through memory and comparing category

¹⁸ At least, we cannot do so based only on the criteria so far laid out to distinguish occurrent thoughts from other psychological states and processes. Remember that the dialectical context here is that of assessing the *empiricist's* account of what sorts of states realize our thoughts. Non-empiricists, e.g., concept rationalists, might want to propose functional role criteria that rule decisively in favor of the non-perceptual vehicles. Perhaps some plausible constraints on concepts entail that perceptual vehicles are unsuitable candidates for realizing the concept role. But we are not attempting here to prejudge the question of whether empiricism or rationalism is the correct account of concepts.

representations, and other directed cognitive activities. Representations in perceptual areas are usually tokened in the course of processing incoming perceptual stimuli. On the sort of model sketched above, though, they are activated in a top-down fashion to assist in these sorts of conceptual processes. That means that there are very likely a set of distinct control states and processes, wherever they may be located in the brain, that are orchestrating this complex activity in perceptual cortices. In order for the appropriate perceptual representations to be retrieved from memory and compared, they need to be content matching with respect to the higher-order representations that activate them. A control system for using perceptual memory in categorization is only effective if higher-order tokenings of DOG lead to retrieval of percepts that also represent dogs. So considering the plausible design features of these mechanisms of categorization and inference, it doesn't seem unreasonable to suppose that they might *also* contain candidate content matching states. And if this is the case, then premise (2) of the empiricist's argument concerning the NCCT can't be sustained.

7. Conclusions

I've argued that the Strong Global version of the perceptual vehicles thesis isn't supported by the neuroscientific evidence. The anti-empiricist case has four main components. First, representations beyond percepts are causally implicated in implementing the causal role distinctive of concepts. Second, evidence of widespread modality-specific cells is either a red herring or threatens to undermine the neat definition of the senses that empiricism requires. Third, falling back on the notion of a 'copy' threatens to either deflate or falsify the thesis. Fourth, appealing to conditions on what makes some neural state a correlate of conceptual thought won't determinately single out activity in perceptual regions as the vehicles of thought. Nothing we now know of allows us to rule out minimal NCCTs that include both perceptual and

non-perceptual regions, and considerations on the design of categorization mechanisms hints that this is in fact a plausible arrangement.

So despite this criticism of SGE, I haven't been arguing decisively for concept rationalism here. In effect, SGE attempts to localize occurrent thoughts in our sensory systems by pinpointing activity in those regions of the brain as the unique filler of the causal role played by concepts. However, occurrent perceptual representations are causally intertwined with activity in numerous non-perceptual regions. This is even more the case when perceptual representations are being activated *as* concepts, under the organism's own control. Undeniably, perceptual representations and processes are tapped as resources in a surprisingly large range of tasks. Some of these are even tasks that we have no *prima facie* reason to think will implicate perception. But this isn't enough to show that thinking just is playing with percepts.

What, then, is the status of empiricist claims that rest on neuroscientific foundations? Ruling out SGE leaves Weak Global Empiricism and Strong and Weak Local Empiricism as open possibilities. The argument of Section 6, however, seems to threaten *any* Strong version of empiricism, since it calls into question our ability to single out perceptual representations as the fillers of the concept role. Given the redundant and distributed way in which intentional content can potentially be represented in the brain, it might be that uniquely perceptual correlates of thought are difficult to find. If this is correct, then the most promising remaining options are Weak Global and Local Empiricism. On both views, at least some thoughts are partially perceptual and partially amodal. This sort of position, somewhere between empiricism and rationalism, is less radical but a better fit for the evidence. Developing specific, testable models that conform to these positions is a task for further research.

Acknowledgements

Thanks to Martin Hahn, Edouard Machery, Pete Mandik, Chase Wrenn, two anonymous referees for this journal, and the editor of this special issue, Rocco Gennaro, for their helpful comments on this paper and earlier versions thereof.

References

- Barsalou, L. W. (1999), 'Perceptual symbol systems', *Behavioral and Brain Sciences*, **22**, pp. 577-609.
- Block, N. (2005), 'Two neural correlates of consciousness', *Trends in Cognitive Science*, **9**, pp. 46-52.
- Caramazza, A., & Mahon, B. Z. (2003), 'The organization of conceptual knowledge: The evidence from category-specific semantic deficits', *Trends in Cognitive Science*, **7**, pp. 354-61.
- Chalmers, D. J. (2000), 'What is a neural correlate of consciousness?' in *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, ed. T. Metzinger (Cambridge, MA: MIT Press).
- Cummins, R. (1989), 'Inexplicit information', in *The Representation of Knowledge and Belief*, eds. M. Brand & R. Harnish (Tuscon, AZ: University of Arizona Press).
- Damasio, A. R. (1989), 'Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition', *Cognition*, **33**, pp. 25-62.
- Damasio, A. R., & Damasio, H. (1994), 'Cortical systems for retrieval of concrete knowledge: The convergence zone framework', in *Large-scale neuronal theories of the brain*, ed. C. Koch (Cambridge, MA: MIT Press).
- Damasio, H., Tranel, D., Grabowski, T., Adolphs, R., & Damasio, A. (2004), 'Neural systems behind word and concept retrieval', *Cognition*, **92**, pp. 179-229.

- Dretske, F. I. (1981), *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press).
- Duncan, J. (2001), 'An adaptive coding model of neural function in prefrontal cortex', *Nature Reviews Neuroscience*, **2**, pp. 820-9.
- Fodor, J. (1975), *The Language of Thought* (Cambridge, MA: Harvard University Press).
- Fodor, J. (1990), 'A theory of content II', in *A Theory of Content and Other Essays* (Cambridge, MA: MIT Press).
- Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001), 'Categorical representation of visual stimuli in the primate prefrontal cortex', *Science*, **291**, pp. 312-6.
- Ghazanfar, A. A., & Schroeder, C. E. (2006), 'Is neocortex essentially multisensory?' *Trends in Cognitive Science*, **10**, pp. 278-85.
- Glenberg, A. M., & Kaschak, M. P. (2002), 'Grounding language in action', *Psychonomic Bulletin & Review*, **9**, pp. 558-65.
- Glenberg, A. M., & Robertson, D. A. (2000), 'Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning', *Journal of Memory and Language*, **43**, pp. 379-401.
- Goldstone, R., & Barsalou, L. W. (1998), 'Reuniting perception and conception', *Cognition*, **65**, pp. 231-62.
- Haugeland, J. (1991), 'Representational genera', in *Philosophy and Connectionist Theory*, eds. W. Ramsey, S. Stich & D. Rumelhart (Hillsdale, NJ: Lawrence Erlbaum).
- Humphreys, G. W., & Forde, E. M. (2001), 'Hierarchies, similarity, and interactivity in object recognition: 'Category-specific' neuropsychological deficits', *Behavioral and Brain Sciences*, **24**, pp. 453-509.
- Kaufer, D. I., & Lewis, D. A. (1999), 'Frontal lobe anatomy and cortical connectivity', in *The Human Frontal Lobes: Functions and Disorders*, eds. B. L. Miller & J. L. Cummings (New York: Guilford Press).

- Kosslyn, S. M. (1994), *Image and Brain: The Resolution of the Imagery Debate* (Cambridge, MA: MIT Press).
- Lycan, W. (1981), 'Form, function, and feel', *Journal of Philosophy*, **78**, pp. 24-50.
- Machamer, P. K., Darden, L., & Craver, C. F. (2000), 'Thinking about mechanisms', *Philosophy of Science*, **67**, pp. 1-25.
- Meltzoff, A. N., & Borton, R. W. (1979), 'Intermodal matching by human neonates', *Nature*, **282**, pp. 403-10.
- Middlebrooks, J. C. (2000), 'Cortical representations of auditory space', in *The New Cognitive Neurosciences*, ed. M. S. Gazzaniga (2nd ed.) (Cambridge, MA: MIT Press).
- Miller, E. K., Freedman, D. J., & Wallis, J. D. (2002), 'The prefrontal cortex: Categories, concepts, and cognition', *Philosophical Transactions of the Royal Society of London B*, **357**, pp. 1123-36.
- Milner, A. D., & Goodale, M. A. (1995), *The Visual Brain in Action* (Oxford: Oxford University Press).
- Murphy, G. (2002), *The Big Book of Concepts* (Cambridge, MA: MIT Press).
- Noë, A., & Thompson, E. (2004), 'Are there neural correlates of consciousness?' *Journal of Consciousness Studies*, **11**, pp. 3-28.
- Polger, T. (2004), *Natural Minds* (Cambridge, MA: MIT Press).
- Prinz, J. (2002), *Furnishing the Mind* (Cambridge, MA: MIT Press).
- Prinz, J. (2006), 'Beyond appearances: The content of sensation and perception', in *Perceptual Experience*, eds. T. S. Gendler & J. Hawthorne (Oxford: Oxford University Press).
- Ramnani, N., & Owen, A. M. (2004), 'Anterior prefrontal cortex: Insights into function from anatomy and neuroimaging', *Nature Reviews Neuroscience*, **5**, pp. 184-94.

Rao, S. C., Rainer, G., & Miller, E. K. (1997), 'Integration of what and where in the primate prefrontal cortex', *Science*, **276**, pp. 821-4.

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000), 'The prefrontal cortex: Response selection or maintenance within working memory?' *Science*, **288**, pp. 1656-60.

Shoemaker, S. (1981), 'Some varieties of functionalism', *Philosophical Topics*, **12**, pp. 83-118.