The Return of Concept Empiricism

Jesse J. Prinz
jesse@subcortex.com

Department of Philosophy, University of North Carolina at Chapel Hill

[Penultimate draft of chapter in H. Cohen and C. Leferbvre (Eds.) *Categorization and Cognitive Science*, Elsevier (forthcoming).

ABSTRACT:
In this chapter, I outline and defend a version of concept empiricism. The theory has four central tenets: Concepts represent categories by reliable causal relations to category instances; conceptual representations of category vary from occasion to occasion; these representations are perceptually based; and these representations are all learned, not innate. The last two tenets on this list have been central to empiricism historically, and the first two have been developed in more recent years. I look at each in turn, and then I discuss the most obvious objection to empiricism. According to that objection, some concepts cannot be perceptually based because they represent things that are abstract, and hence unperceivable. I discuss two standard examples: democracy and moral badness. I argue that both can be explained using resources available to the empiricist.

The history of Western philosophy can be viewed as a debate between rationalists and empiricists. Rationalists emphasize innate concepts, the power of *a priori* reasoning, and the unreliability of perception. Empiricists regard perception as the source of our concepts and the primary means of attaining knowledge. Since Plato and Aristotle, the pendulum has been swinging back and forth between these positions. The high point in this debate occurred in the 17th and 18th centuries, when continental rationalists, such as Descartes and Leibniz, revamped rationalism, and British Empiricists, such as Lock and Hume, worked out the empiricist alternative. Cognitive science was born as the pendulum turned back towards rationalism. Before the cognitive revolution, Skinner had allied himself with the empiricists, and Chomsky launched a self-consciously rationalist assault on Skinner's research program. Within psychology, the rationalist trend has continued. For example, a sizable percentage of research on concept acquisition focuses on innate domains of knowledge. There are, however, also signs of dissent. Connectionism and situated cognition have attracted considerable interest over the last decade, and both approaches depart from rationalism in significant way. There has also been a revival or more traditionally empiricist theories. Of special interest is the work of Larry Barsalou and his colleagues. His group has, more than any other, carried the torch for the likes of Locke and Hume. Barsalou's work has been my point of departure.

In this chapter, I argue that a traditional brand of empiricism has a lot to offer. I think it is time for the pendulum to swing back. I will begin my case by presenting the core components of the empiricist theory that I favor. The most central tenet is that concepts have the basis in perception. A second tenet, which was equally important to Locke and Hume, is that concepts are learned. To these classical tenets I add two more contemporary suggests: concepts refer to categories in the world via reliably causal relations and concepts are contextually variable. I will discuss these four tenets in reverse order. My discussion of the first two is primarily directed at philosophers. My discussion of the second two will be of equal concern to psychologists, who may wish to skip ahead. Together, I think these four tenets add op to a defensible version of empiricism. There is, however, one serious objection, stemming from the tenet that concepts are perceptually based. This is plausible in the case of concrete categories, but we are also capable of thinking about abstract things. It is not clear how perceptually based concepts can handle abstract thought. I conclude with a discussion of this objection.

1. Concept Empiricism

1.1 Representing and Doing: Two Faces of Concepts

In a recent paper, Jerry Fodor (2004) has suggested that rationalists about concepts can be distinguished from defenders of other theories by their view of what concepts are for. Rationalists claim that concepts are for thinking. More precisely, to have a concept is to be able to think about something. Having the concept DOG is being able to think about dogs. A better word for thinking, in this context, might be representing. Concepts are primarily in the business of representing. For opponents of rationalism, including empiricists, having a concept is being able to do something. For example, having the concept DOG might be construed as the ability to categorize or interact with dogs. Fodor's way of setting things up distinguishes two broad functions for concepts: rationalists say that concepts are primarily in the business of *representing*, and opponents or rationalism say that concepts are primarily in the business of *doing*. This distinction should not be regarded as a disjoint dichotomy. Empiricists do not deny that concepts representing. Rather, they claim that concepts have other equally important functions. Empiricists say that concepts must be able to representing things in a way that facilitates interaction with those things. Representing must be in the service of doing.

The difference between Fodor and the empiricists is captured by a difference in the nature of the mental representations that they postulate. Fodor thinks concepts are like words. They are arbitrary symbols in a language of thought. There is little one can do with an arbitrary symbol. A symbol does not include any instructions for how to interact with the category that it represents. For the empiricist, concepts are more like mental images, or inner models. Representations of that kind can be used to guide action. We can read features of a category off of our concepts if empiricism is right.

I will discuss these issues of representational format below. In this section, I want to consider another question about the distinction between representing and doing. Fodor attempts to separate these two functions by developing a theory of representation that would allow concepts to represent things without encoding the kinds of features that would allow them to do anything. That theory constitutes one of the best current explanations of how concepts represent. If the theory supports Fodor's rationalism, then empiricists have reason for concern. My goal here is to show that Fodor's theory of reference is actually better suited for empiricism.

Fodor (1990) develops a theory of representation that promises to explain how concepts represent without making any mention of what they do. This allows him to secure the

conclusion that representing is prior to doing, which is a central tenet of rationalism. I begin with a summary of the theory. According to Fodor, a concept represents that which would reliably cause the concept to be activated. A concept represents dogs if encounters with dogs would ordinarily cause that concept to activate. This is only a first approximation. Formulated in this way, the account faces an obvious objection. Our DOG concepts are activated when we encounter dogs, but they are also activated when we encounter things that merely look like dogs, e.g., foxes in bad lighting. If concepts represent *anything* that activates them, any concept that represents dogs would also represent foxes. To solve this problem, Fodor notes that there is an asymmetric dependency relation between dogs and foxes with respect to our DOG concepts: foxes would not cause our DOG concepts to activate were it not for the fact that dogs do, but the converse is not true: the fact that dogs cause our DOG concepts to activate is not a consequence of the fact that foxes do. Fodor construes this asymmetry synchronically, in terms of counterfactual dependencies. I will not argue the point here, but I think the best way to make sense of it, is diachronically. A DOG concept is one that was created in the context of dog encounters. Fox encounters would not cause the concept to activate were it not for dogs having done so *in the past*, but not conversely. On this reading, a concept represents a category when two conditions are met:

> *Nomological causation*: the concepts is disposed to be reliably activated by encounters with members of the category, and
>
> *Etiological causation*: encounters with members of the category played a role in the acquisition of the concept.

Fodor thinks that concepts represent in roughly this way (with a synchronic condition in place of the etiological causation clause). He also thinks that this story favors the hypothesis that concepts are primary in the business of representing, not doing. To see why, it is important to consider two other alternatives to this causal theory of reference. According to one view, concepts refer by resemblance. They are mental images that are structurally isomorphic with the things they represent. According to another view concepts are feature sets that refer via description. The concept DOG refers to dogs, because the concept dog contains a collection of features describing dogs, and dogs are the only things that satisfy the description. DOG contains FURRY, BARKS, QUADRUPEDAL, and so on. By denying these two theories of reference, Fodor is able to defend the view that concepts are unstructured arbitrary symbols (Fodor, 1998). They are words in a language of thought. Fodor can defend the language of thought story only be arguing that concepts do not depend on description or resemblance to refer. An individual word does not describe anything, and it does not look like what it refers to. By embracing a causal theory of reference, Fodor explains how word-like mental representations can refer. Without this, it would be difficult to maintain that concepts are couched in an arbitrary code. It would also be hard to maintain that concepts are primarily in the business of representing. An arbitrary symbol cannot be used, on its own to recognize dogs or draw inferences about dogs. It is a dog symbol in the purest sense: it represents dogs and does nothing else. A mental image of a dog represents dogs and can also be used to recognize them. A dog description represents dogs and can also be used to draw inferences about them. Fodor's causal theory of reference secures his hypothesis that concepts are primarily in the business of representing, not doing.

Fodor's mental word theory is radically different from the way most psychologists think about concepts. Psychologists emphasize the role that concepts play in categorization. If concepts are tools for categorizing, they cannot be unstructured word-like entities. They must be built up from features. Some psychologists say that DOG is a prototype; others say it is a mini-theory; and still others say it is a set of exemplar representations. As a rationalist, Fodor thinks that a theory of concepts need not explain how we categorize. His mental word theory is ideally suited for rationalism. After all, words in public languages represent, but we cannot categorize with them; they are arbitrary symbols. On Fodor's view categorization is achieved by independent mechanisms. He doesn't offer an account, but he might say that we have complex mental databases containing perceptual information, theories, prototypes, memory traces, and any number of features and facts. That is to say, categorization is achieved using the kinds of mental mechanisms that psychologists postulate. Think of a concept as a label on a large mental file. Information in that file, and information stored elsewhere can play a role in categorization. The concept hovers safely above the overflowing sheets and scraps in the file. Items in the file represent what category members look like, the ontological domain they belong to, the attributes of specific instances, and so forth. But only the label represents the category itself.

On the face of it, Fodor seems to have what he wants. He has a theory of how mental representation works that is consistent with rationalism. Concepts are arbitrary symbols that can be used for nothing other than representing categories. They cannot be used to draw inferences, to plan actions, or to categorize. All of those functions are handled by the contents of our mental files. But this picture is very odd. It renders concepts needlessly anemic. Why should we say that concepts are arbitrary labels, rather than identifying concepts with the contents of our mental files? After all, the contents of those files do much more work. They allow us to categorize and act. Moreover, these files are absolutely essential for Fodor's own theory of representation. A mental label represents a category by being reliably activated by instances of that category. But the label can be activated by category instances only if we have mechanisms that allow us to recognize those instances. An arbitrary DOG symbol can be triggered by dogs only if we have resources for recognizing dogs. Fodor assumes that all of the necessary resources are contained in our mental files. But once he makes that concession, the arbitrary labels begin to look unnecessary. It seems we should identity concepts with the file contents, rather than the file labels. We should say that concepts are the mechanisms that allow us to recognize categories rather than arbitrary mental words that flash on in the head when a category has been recognized. The labels are entirely unnecessary.

The moral is that Fodor's theory of representation may not favor rationalism after all. Once we adopt a causal theory, we are forced to postulate mechanisms that allow us to reliably detect category instances. Once we postulate such mechanisms, we might as well identify them with concepts. We do not need to postulate arbitrary labels. Concepts can be complex databases. Such databases allow us to represent, but they also allow us to do things; they allow us to interact successfully with the world. So representing and doing are not disjoint functions, on this picture. They are intimately linked. On Fodor's theory, we think using unstructured symbols. DOG is just an arbitrary word in the language of thought. On the view I am recommending, DOG is constituted instead by the representations used to identify dog. Thus, DOG is constituted by features that tell us what dogs look like and what their behavioral dispositions and affordances for interaction are. These features allow us to represent dogs, by securing reliable casual relations with then, but they also allow us to recognize dogs, and do things with dogs.

Rather than saying that concepts are for representing, I would say that representations are for doing. The only reason we represent the world is to make our way through it. If concepts are not guides to possibilities for action, they are not useful. Concepts that merely represent belong to the fictional realm of pure Cartesian egos. Conceptual capacities that evolved in the real world allow us to run for cover or play fetch. But concepts also represent. The mechanisms that allow us to identify objects and interact with them also, thereby, establish reliably causal relations with those objects. Fodor himself shows how such causal relations can be used to establish reference. Ironically, his theory of reference fits perfectly with the anti-rationalist program.

1.2 Variable Mechanisms

Fodor has reasons for identifying concepts with arbitrary labels, rather than the contents of mental files. He did not come to this conclusion without any arguments. He thinks that all theories that identify concepts with complex data structures, rather than unstructured word-like entities, are hopeless. If he is right, concepts cannot be identified with the complex mechanisms in our mental files that we use to identify members of a category. He has three main reasons for this conclusion:

1. The mechanisms in question do not combine compositionally. Concepts must be compositional, because that is the only way to explain our capacity to continually generate new thoughts. Therefore concepts cannot be the mechanisms of categorization.
2. The mechanisms in question are highly varied and highly variable. Thus they are unlikely to be the same from person to person or time to time. Concepts are also used to assign meanings to words. People must assign the same meanings. Otherwise they would not be able to communicate. So concepts cannot be so variable, and they cannot be the mechanisms of categorization.
3. Concepts represent via reliable causal correlations with the things they represent. If concepts were as variable as the mechanisms that we use to identify category instances, then they would not be highly correlated with the categories they represent. Different concepts would be used for different encounters with the same category. That would prevent concepts from serving as representations of those categories. Therefore, concepts cannot be mechanisms of categorization.

I discuss the first two problems elsewhere (Prinz, 2002; Prinz and Clark, 2004). I want to turn my attention to the third—the claim that representation requires conceptual invariance. This claim is only implicit in Fodor, but I think it is the best argument for the view that concepts are like words.

Let us assume, with Fodor, that categorization cannot be achieved using invariant representations. We have mental files filled with different kinds of representations that are recruited on different occasions. We might even generate new representations for a category on the fly, by combining information in the file for that category with other information that pertains to the present context. The question is, Can such varied representations represent the category, as opposed to representing transient features of the category (e.g., its appearance in this instance right now)? I think the answer is yes. To make this point, I need to show that the theory of representation sketched earlier is applicable to variable representations. If variable

representations satisfy the two conditions in that theory (etiological and nomological causation), then variability presents no barrier to explaining how concepts represent.

Let's begin with the etiological condition. According to that condition a mental representation can represent a category only if the representation is an instance of a type of representation that was acquired as the result of an encounter with that category. This condition is satisfied by the representations that we use in categorization even if those representations are variable. The highly varied representations used to categorize instances of a particular category have something in common: they derive from the same mental file. That file was established when we first encountered instances of the category in question. Every time we encounter an object that matches information already contained in the file, we have an opportunity to add new information. We can also expand files by reflecting on the information that they already contain. Because this information is bundled together in memory, we are able to do the requisite bookkeeping. If we represent a category using item A from a file on one occasion, and item B from the same file on another occasion, we know that A and B are culled from the same source. In this sense, the heterogeneous representations used to represent a category on different occasions satisfy the etiological causation condition on representation, which I introduced earlier. All of the items in a mental file trace back to an initial time when the file was created, and none of the items in that file would be there if it were not for the initial representations formed when we first encountered an instance of the category. Suppose my DOG file was created when I first saw a dog, and suppose that dog was a golden retriever. Later in life, I see a Pomeranian for the first time, and it looks similar to the retriever in certain respects, so it ends up in my DOG file. Now, one afternoon, I see a fox, which looks a lot like a Pomeranian, so it triggers the Pomeranian representation. Does this entail that my DOG file represents dogs *and foxes*? No. When I call a fox a dog, I am making a mistake. The DOG file was created as the result of an encounter with a dog. Other animals may happen to activate items in the file, but, since the file itself traces back to dogs, these are cases of misrepresentation. I conclude that variable representations can satisfy the etiological condition on representation.

Variable representations also satisfy the nomological causation condition, at least when they are considered collectively. There is a reliable causal relationship between encounters with members of a category and representations derived from the file for that category. My Pomeranian representation is not reliably caused by dogs, in general. It might not be activated when I encounter a sheep dog. But my Pomeranian representation is a member of a mental file containing variable dog representations, and these collectively are reliably caused by dogs. In other words, dog encounters reliably cause us to access the dog file. Items in the dog file are, in that sense, under the nomological control of dogs. Items in a mental file can be said to refer to the category that the file reliably detects. Thus, the nomological condition on reference can be met, even if the representations we use to categorize dogs are highly variable. I conclude that variability poses no barrier to representation. Concepts can represent categories even if they are instantiated in a variety of different ways.

On the view I am considering, the same concept will be constituted by different representations on different occasions. When you see a small long-haired dog, you will use one dog representation, and when you see a large short-haired dog, you may use another. In this sense, concepts vary with context. Barsalou (1987) reviews evidence that favors this conclusion. There is reason to think that we represent the same category in a variety of different ways. A typical dog for in a French restaurant will differ from a typical dog in the arctic tundra. Barsalou shows that judgments of typicality vary as a function of imagined contexts. He concludes that

concepts are not small sets of fixed features, much less unstructured words in a language of thought. Rather concepts are temporary and variable constructions in working memory. They are drawn up in task-sensitive ways from large data structures in long-term memory.

1.3 Perceptual Vehicles

The story, thus far, has two core tenets. One, borrowed from Fodor, is that concepts are representations of categories, and they represent, in part, by being reliably caused by category instances. The other, borrowed from Barsalou, is that concepts are highly variable constructions in working memory. The two tenets go together naturally. The variability ensures that concepts can be reliably activated by encounters with category instances. There is a third core tenet of the view that I favor, which is also found in Barsalou (1999). Concepts are perceptually based.

To say that concepts are perceptually based is to say that they are made up from representations that are indigenous to the senses. Concepts are not couched in an amodal code. Their features are visual, auditory, olfactory, motoric, and and so on. They are multimedia presentations. This tenet lies as the heart of classical British Empiricism. Hume (1739) says, "All our ideas are nothing but copies of our impressions." I call in the modal specificity hypothesis.

The evidence for modal specificity comes from a variety of sources. Barsalou (1999) summarizes findings from psychology, neuroscience, and linguistics that are consistent with the idea that we think in perceptual codes. Damasio (1989) has argued, on the basis of functional neuroanatomy and neurological deficits, that thinking involves reactivating the perception centers that would be active if we were perceiving the things we were thinking about. Barsalou et al. (1999) have shown that people spontaneously use imagery in cognitive tasks. Lakoff (1987) has shown that there are pervasive uses of perceptual metaphors when we verbally describe abstract domains.

Barsalou (this volume) offers a up-to-date review of findings that lend support to the claim that concepts are perceptually based. I will not repeat his survey here, but let me briefly mention two of representative results. First, consider a study by Borghi et al. (in press). If concepts are constituted by amodal symbols, then the features related to a concept should be organized as a list or semantic network. Proximity in a semantic network should be based on semantic relatedness (e.g., dimensions and attributes should be directly linked) or strength of association (co-instantiated lexical items should be linked). If concepts are constituted by perceptual representations then further factors will contribute to feature organization, such as salience and special proximity. The perceptual theory of concepts predicts that features that are close to each other on a category instance will be close to each other in our mental representation of that category instance. Borghi et al. tested for this by giving subjects a property verification task. Subjects first read sentences such as "you are washing a car." Then they had to answer questions such as "do cars have trunks" or "do cars have steering wheels." Both features are strongly associated with cars, but subjects who were given the sentence about car washing were faster to answer the question about trunks. These results reversed when subjects began with the sentence "you are driving a car." These results show two things. First, they re-confirm that representations are variable. The way we think about cars depends on the context (driving vs. washing). Second, our representations of categories are spatially organized; parts that are farther from our current perspective take more time to access, even if those parts are strongly associated with the category. This is predicted by the perceptually based theory of concepts and not by the amodal theory.

As a second example, consider a recent study by Pecher et al. (2003). They asked subjects to answer a series of property verification sentences. For example, subjects might be asked: Are blenders loud? Are cranberries tart? Do leaves rustle? The questions involved familiar features of objects. If concepts were represented in an amodal code, then features such as "loud," "rustle," and "tart" would be represented in a similar way. If concepts are represented using modality specific features, then the first two are stored in one sensory modality (audition) and the other is represented in another sensory modality (gustation). Thus, the perceptually based theory predicts that when subjects move from a question about the loud blenders to a question about rustling leaves, they should be faster than if they had to move from a question about loud blenders to a question about tart cranberries. Shifting modalities should incur switching costs, if concepts are coded in a modality specific way. This is just what Pecher et al. find. The task shows not only that perceptual features are associated with our concepts, but that we use such features when performing conceptual tasks. Defenders of amodal features do not predict this.

These findings provide empirical support for the hypothesis that concepts are perceptually based. Further support comes from theoretical considerations. Above, I endorsed the view that concepts represent by being reliably caused by category instances. This theory of representation is very popular in philosophical circles, and it has been defended by rationalists, such as Fodor. But the theory can actually be used to argue for the modality specificity of thought. If concepts are reliably caused by the categories that they represent, then every concept must be associated with a collection of perceptual features. Objects out there in the world can cause mental events only by impinging on our senses. So anyone who buys into the theory of representation presented earlier is committed to the view that we can perceptually identify category instances and that doing so is essential to representation. If perceptual states are essential for getting concepts to represent, we can simply hypothesize that concepts are copies of those perceptual states. This is just a minor addition to an argument already presented. I argued that concepts are the mechanisms by which we categorize. To that, I now add that those mechanisms must be realizable in perceptual media, because categorization requires identification of *perceived* objects. Empiricism sounds radical at first, but it is actually consistent with the simple and obvious point that perception is needed to apply our concepts to things in the world. Every serious theory of concepts is committed to that. Empiricists take this universally accepted principle and runs with it. If concepts are associated with perceptual representations, perhaps that's all we need. Postulating a further class of representations (amodal symbols) is unnecessary.

The modal specificity hypothesis is compatible with a parsimonious theory of how the brain evolved (see also Barsalou, 1999; Churchland, 1986). At first, creatures were input-output machines. The world causes sensory stimulation, and sensory stimulation caused programmed responses. Consider how flies avoid swatters, and you'll get the idea. Over time, creatures evolved the capacity to store perceptual records of objects that they had encountered in the past, as well as the past consequences of those encounters. This allowed for much creative flexibility of response. Such creatures can respond differently to different objects even if they were not hardwired to recognize those objects. Finally, creatures evolved the capacity to re-activate the stored perceptual records in the absence of sensory stimulation, and the capacity to manipulate those records in working memory. That's what we do, and we do it better than any other creature on Earth. This story does not require the evolution of any special amodal codes.

1.4 Innateness

The theory of concepts that I favor has one more tenet, which also derives from the empiricist tradition. Locke (1690) began his case for empiricism by arguing against the doctrine of innate ideas. He believed that there were no innate concepts or principles. Empiricism was motivated by the idea that the attainment of concepts must involve learning, rather than triggering innate knowledge. These days, nativism about concepts is very popular. The majority of researchers working on concepts believe that we have quite a bit of innate machinery. The most popular suggestion is that we have innate knowledge of basic ontological domains, such as macro-object physics, biology, and psychology.

I don't believe that any of these domains is innate. That is to say, I do not think we have innate domain-specific knowledge that contributes to structuring our concepts. I cannot adequately defend the claim here. What I will offer instead is a few brief remarks and pointers. I hope that such an abbreviated discussion can at least serve to motivate hard reflection on the widespread nativist dogma. The innateness of core domains remains an open question for research.

First consider folk physics. Spelke (1994) argues that our understanding of certain physical principles must be innate because we seem to understand these principles early, and other principles, which are more perceptual salient, are not understood as early. For example, three-month olds seem to understand a principle of cohesion: objects move as connected wholes. When they habituate to what looks like a bar moving behind an occluder, they dishabituate if the occluder is removed to reveal two bars moving in sync. They dishabituate less if the bar behind the occluder is solid, consistent with the adult expectation (Kellman and Spelke, 1983). In contrast, infants at the same age do not understand gravity. They do not expect an object teetering far off the edge of a supporting surface to fall (Needham and Baillargeon, 1993). Spelke thinks gravity is more obvious perceptually than object cohesion, so the latter must be innate.

Several things can be noted in response. First, expectations of cohesion can be learned by observation. When we see two shapes moving along the same spatiotemporal trajectory, they are almost always connected. Slater et al. (1990) showed that newborns do not make the coherent object assumption. Second, even cohesion were not learned, there is little reason to think it is a conceptual capacity, rather than a feature of how our perceptual or attentional systems pick out objects (Scholl and Leslie, 1999). Third, gravity is often violated in an infant's word, so it is unsurprising that infants are slower to learn about it. Infants see hanging mobiles and doorknobs (Baillargeon et al., 1995), and they have comparatively little experience with their own bodies falling. Finally, the Kellman and Spelke results may be an artifact of display complexity (compare Bogartz et al. 1997). It is widely known that infants will, at baseline, often stare at two objects for a longer time than they will stare at one object. In the Kellman and Spelke study, infants stare longer when an occluder is removed to reveal two bars rather than one. Perhaps they are staring longer simply because two objects are more interesting than one. Despite a variety of clever control conditions, Kellman and Spelke do not adequately rule out this hypothesis. For example, in one control condition, they habituate infants to a pair of moving bars with no occluder. At test, infants stare longer at one bar rather than two. This shows that infants do not always stare longer at two bars. If the infants construed the display in the first experiement—the one with the occluder—as two bars rather than one, they should get bored of seeing two bars and show excitement when the occluder is removed to reveal a single bar. Since infants in that experiment show excitement about the two bars Kellman and Spelke conclude that

infants construed the habituation display as a single bar. But this conclusion is too hasty. In the control condition, infants see two bars moving with a gap in between them. This looks just like the two bars in the test condition. In the version with the occluder, infants do not see two bars with an open gap between them. The occluder fills the gap. In other words, the appearance of the bars in the control condition is just like the appearance of the bars in the test condition, because there is an unfilled gap. In the occluder condition, the appearance changes, because a gap appears, where there had been a surface. This difference may be enough to erase the effects of habituation and re-engage the infants attention. This suggestion is highly speculative, of course, but it is intended as a reminder that it is difficult to infer what infants are thinking from their performance in studies of this kind. When it looks like infants understand a feature of the physical world, they may actually be showing their preference for perceptual similarity.

Now consider folk biology. Keil (1989) found that preschoolers think that an artifact cannot be turned into a living thing even if its appearance is altered to look just like a living thing, and conversely. But preschoolers do think that appearances can change one living thing or one artifact into another. Within domain transformations affect identity, but cross domain transformations do not. This suggests that preschoolers distinguish artifacts from living things. He speculates that they are innately sensitive to this distinction.

The fact that children think that within-domain transformations can affect identity is hardly surprising. They know about many cases where lexical labels change with appearances: caterpillars become butterflies, boys become men, seeds become trees, construction paper becomes an art project. But children do not experience many cross-domain transformations, and they may be explicitly taught that some of these are impossible. When playing with dolls, for example, parents may say, "That's not a real baby; it's make believe." Moreover, artifacts and living things are very easy to distinguish perceptually. Living things have faces, and they tend to be fuzzy, irregular in their movements, symmetric along a central access. The fact that these features co-occur may lead to an early, robust, perceptually learned category. By kindergarten, children have been told stories about where entities in this category come from, and they may be reluctant to accept that artifacts can gain admittance by mere change in appearance. I bet answers would differ if they heard about an animal-like robot coming from Mommy's tummy.

Finally, consider folk psychology. Kids attribute mental states to others. No one knows exactly how this "mindreading" ability comes about. It has a relatively fixed time-course in development, with the capacity to attribute beliefs emerging after the capacity to attribute desires and perceptions. Mindreading seems to be lacking in apes (Povinelli and Eddy, 1996; though see Hare, Call, & Tomasello, 2001), and it is selectively impaired in autism (Baron-Cohen, 1995). These findings indicate an innate domain. Or do they?

Mindreading abilities may originate in our capacity to imagine things (Gordon, 1986; Goldman, 1989). We can imagine situations that are not actually occurring. This is part of our general executive working memory capacity. Imagining others is, at first, like imaging ourselves in another situation. Empathy, emotional contagion, tracking eye-gaze and other relatively primitive, pre-conceptual abilities may orient us towards others in a way that promotes taking their perspective. If apes lack mindreading abilities, that may be due to a more general limitation in their capacity to imagine. Apes may have a less developed executive working memory. That would prevent them from actively choosing to imagine what things are like from another perspective.

The sequence of mindreading development in human children is hardly surprising. Desires and perception are usually directed towards perceptually present objects; beliefs are not.

Desires and perceptions are associated with characteristic conscious experiences; beliefs are not. Desires and expressions can be attributed using simple accusative verb constructions; belief attributions have sentential clauses as direct object. All these things make desire and perception easier to learn. Moreover, even if the sequence of learning is fixed, the time course is not. Those who think mind-reading is innate emphasize the fact that children begin to attribute beliefs around the age of 4. It turns out that this age varies across cultures. In some cultures, the time-course is much slower (Vinden, 1999). The fact that kids in our culture master belief attribution at 4 may reflect facts about how kids are socialized in the Western world. It may also reflect facts about English and other Western languages. Belief attribution requires embedded clause constructions, which are mastered around the age of 4. In fact, good performance on belief attribution tasks is well correlated with the tendency to interpret that-constructions as complement clauses. In general, degree of social interaction and linguistic skills seem to be the best predictors for mindreading abilities (Garfield, et al., 2001). People with autism may be impaired as a consequence of more general social and linguistic impairments.

These remarks are not intended to account for the wealth of findings in support of innate domains. They merely demonstrate that such findings are open to multiple interpretations. Until we have investigated non-nativist explanations of the developmental evidence, nativism cannot be taken for granted. We certainly have many innate capacities, faculties, and biases, but concept-guiding ontological domains may be learned. If they are, then Locke's empiricists theory of concepts may be closer to the truth than most researchers would dare to imagine.

1.5 Summary

The theory I have been present has three core tenets:

1. Concepts represent categories via nomological and etiological causation
2. Concepts are variable constructions in working-memory
3. Concepts are built of from modality specific memory traces
4. Concepts, and the core domains that organize them, are all learned

2 and 3 are the central features of Barsalou's theory. 3 and 4 are the central features of classical British empiricism. 1 has been a central theme in contemporary philosophy of mind. None of these tenets enjoys widespread support in psychology. The acceptance of any would be a significant departure from the orthodoxy. My goal here has been to show that the orthodoxy may be mistaken. Empiricism is not widely embraced, but it is consistent with current evidence. I would urge researchers to take the empiricist program seriously and test it. If we simply assume that concepts are innate, invariant, and amodal, we may fail to discover important facts about concepts.

If empiricism is a viable theory, why is it so rarely endorsed. One answer is that the majority of people working on concepts in philosophy and psychology believe that there is a fatal objection to empiricism. In particular, many researchers dismiss empiricism on the grounds that it cannot accommodate concepts that represent things that are very abstract. Empiricists claim that concepts are like mental images. That makes sense for concrete categories, but it won't work for many other cases. Therefore, empiricism is, at best, an incomplete theory of concepts. I will briefly address this concern.

3. The Abstract Ideas Objections

The standard knee-jerk reaction to empiricism is that it cannot handle abstract concepts. Perceptually derived concepts are like mixed media images, and no image, no matter how complex can depict such lofty abstractions as TRUTH, MORALITY, and DEMOCRACY. This objection has been discussed in more detail elsewhere (Barsalou, 1999; Prinz, 2002; see also Lakoff, 1997). Barsalou, for example, tells a perceptual story about how we understand TRUTH. Here, I can only gesture at a response to this objection, illustrating with DEMOCRACY and MORALITY.

The first thing to note is that non-empiricitst theories may have no advantage here. The major difference between empiricism and standard non-empiricists theories is that empiricists say concepts are implemented using modality specific codes. If concepts were amodal, we wouldn't face the question of how we can depict democracy, but we would face an equally challenging question. How can an arbitrary amodal symbol inside the head represent democracy? How can it represent anything at all? The appeal to amodal symbols give the illusion of explaining abstract concepts, but it really makes no progress. We still need to explain how a symbol in the head can represent something complex and unperceivable. If concepts represent by their causal relations to the world, the question becomes, how can a symbol in the head get causally connected to democracy if democracy isn't something one can see, taste, or smell?

Defenders of amodal symbols have a standard strategy for dealing with such cases. They say that abstract concepts are understood by complex networks of inferentially related concepts. We know what to infer if we are told that something is a democracy. We know it is system of governance or collective decision-making. We know that, within a democracy, the governed have votes. And so forth.

There are two things to say about this strategy. The first is that it is that it does not require amodal symbols. It just requires symbols. The inferences used to understand what a democracy is can be implemented by a lexical network, relating the English word "democracy" to other words in English (or some other public language). These words are represented as stored perceptions of sounds or marks or gestures. They are perceptually-based, rather than amodal, and the associations between them are learned by listing to discourse about democracy.

This approach to the concept of DEMOCRACY suggests that there is a role for labels after all. But they are not necessarily labels in a language of thought. They can be the labels we devise in public languages. Language can be used, in this way, to expand the power of thinking. Verbal labels serve as placeholders for ideas that are too complex to hold before one's mind all at once. Labels can also facilitate reasoning, by presenting thought is a logic-friendly, linguistic code. If abstract concepts like DEMOCRACY can be adequately explained by networks of labels, then empiricists explain abstract concepts as readily as their opponents.

The second thing to notice about the label strategy is that it cannot be sufficient on its own. Ultimately, these labels must be pinned down in the senses in order to be applied in the world. If we taught a monolingual Mandarin speaker some sentences containing the English word "democracy" she would not thereby know the meaning of the term. To know what the word means, she would have to know where and when to apply it, and she would have to know how to go about making a decision democratically. Mastery of linguistic inferences allows us to reason about democracies, but full mastery of the concept requires some capacity to link it up with the world. The word "vote" may be tied to a stored record of behavioral practices. If someone says, "we are taking a vote" in the right context, we know to raise our hands. In another context, voting involves going to a voting booth and filling out a form. The sheer variety

of practices involved in understanding what a democracy is suggests that we must have a highly variable representation of that category, in accordance with the theory of concepts that I have been defending. Amodal symbols offer no special advantage when explaining these practices. Learning how to vote involve recognizing scenarios and behaving in particular ways. A network of arbitrary symbols, either in a language of thought or in a a public language, can contribute something to our competence, but it cannot be the whole story. When it comes to DEMOCRACY, amodalists have no advantage.

Labels are important for some abstract concepts, but not all. I want to consider a second class of cases: moral concepts (for a detailed discussion, see Prinz, forthcoming). Take the concept MORALLY BAD. On the face of it, this presents a challenge for empiricists. What does moral badness look like? What image could work? No, of course, we can all imagine things that are morally bad, and images of certain naughty scenarios may play a role in understanding moral badness. But that cannot be the whole story. We must be able to think of these scenarios that they are bad. What image allows us to do that?

Hume points us towards an answer to this question. He saw that our understanding of morality is intimately tied to certain emotional reactions. Moral badness is related to a variety of bad feelings. This proposal seems to hold up in recent research. When a stranger does something bad, we feel angry, contemptuous, or disgusted. If the person close to us does something bad, we feel disappointed. If we ourselves do something bad, we feel guilty or ashamed. The concept of MORAL BADNESS is a disposition to experience one of these emotions in response to actions of ourselves and others. The emotion we experience depends on the nature of the action. Emotions obey a systematic logic. Certain emotions arise in certain circumstances. Rozin et al. (1999) have shown that anger pertains to actions that bring harm to persons or property; contempt pertains to actions against the social order; disgust pertains to actions against nature. To think that something is bad is to be disposed to feel the appropriate emotion and to be disposed to feel badly if you or others fail to feel the appropriate emotion.

To reconcile this account of moral concepts with empiricism one needs an empiricist theory of what emotions are. I have defended such a theory elsewhere (Prinz, 2004). The central feature of the account is borrowed from William James (1884). Emotions are perceptions of patterned changes in the body. Emotions get their meaning by being tied to particular kinds of circumstances. Fear represents danger, because it is reliably triggered by dangers. This process need not involve any concepts. A loud noise or a sudden loss of support can trigger fear. Likewise for the "moral emotions." A glare can cause anger. But the causes are often more complex. We may get angry at someone who verbally expresses opposition to democracy, for example. The key point is that emotions themselves are perceptual states, and they derive their meaning, like any mental representation from the kinds of things that cause them to activate. Emotions are easy to accommodate within an empiricist framework, and they extend the range of concepts that we can grasp (see also Barsalou, 1999). Some of our loftiest concepts may be our most visceral, and no appeal to amodal symbols can capture the motivational tug and push of moral thinking.

These two strategies, verbal and emotional grounding, do not exhaust the options that are available to the empiricist. Other concepts may be handled in other ways. We need to reflect hard about abstract concepts and investigate each individually before condemning the empiricist. In each case, I predict, the empiricist will have sufficient resources. Recognizing this is a crucial step in resuscitating empiricism, because most cognitive scientists assume that abstract concepts

pose a fatal objection. That assumption is rarely defended, and it may collapse under scrutiny. If it does, conceptual empiricism may regain the centrality that it once enjoyed.

References

Baillargeon, R., L. Kotovsky, and A. Needham (1995). The acquisition of physical knowledge in infancy. In D. Sperber, D. Premack, and A. J. Premack, eds., *Causal cognition: A multidisciplinary debate*. New York: Oxford University Press.

Baron-Cohen, Simon (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Barsalou, L. W. (1987). The Instability of Graded Structure: Implications for the Nature of Concepts. In U. Neisser, ed., *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge: Cambridge University Press.

Barsalou, L. W. (1993). Flexibility, Structure, and Linguistic Vagary in Concepts: Manifestations of a Compositional System of Perceptual Symbols. In A. Collins, S. Gathercole, M. Conway, and P. Morris, eds., *Theories of Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barsalou, L. W. (1999). Perceptual Symbol Systems. *Behavioral & Brain Sciences*, 22, 577-660.

Barsalou, L. W., K. O. Solomon, and L. L. Wu, L.-L. (1999). Perceptual Simulation in Conceptual Tasks. In M. K. Hiraga, C. Sinha, and S. Wilcox, ed., *Cultural, Typological, and Psychological Perspectives in Cognitive Linguistics: The Proceedings of the 4th Conference of the International Cognitive Linguistics Association, Vol. 3.* Amsterdam: John Benjamins.

Bogartz, R.S., Shinskey, J.L., & Speaker, C.J. (1997). Interpreting infant looking: The event set x event set design. *Developmental psychology*, 33, 408-422.

Borghi, A.M., Glenberg, A., Kaschak, M. (in press). Putting words in perspective. *Memory and Cognition.*

Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.

Cognition and Emotion, 13, 19-48.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition, 33,* 25-62.

Fodor, J. A. (1990). *A theory of content*. Cambridge, MA: MIT Press.

Fodor, J. A. (1998). *Concepts: Where cognitive science when wrong.* Oxford: Oxford University Press.

Fodor, J. A. (2004). Having concepts: A brief refutation of the 20th century. *Mind and Language*, 19, 29-47.

Garfield, J. L., Peterson, C. C., Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind & Language*, 16, 494-541.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, 4, 161-85.

Gordon, R. (1986). Folk psychology as simulation. *Mind and Language* 1, 158-

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour, 61*, 139-151.

Hollebrandse, B. (2003). Long distance Wh-extraction revisited. *Proceedings of the Boston University Conference on Language Development*.

Hume, D. (1739/1978). *A treatise of human nature*. Nidditch, P. H., ed. Oxford: Oxford University Press.

James, W. (1884). What is an emotion? *Mind*, 9, 188-205.
Keil, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
Kellman, P. J., and Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*, 483-524.
Lakoff, G. (1987). *Women, fire, and dangerous things*. Chicago, IL: University of Chicago Press.
Locke, J. (1690/1979). *An essay concerning human understanding*. P. H. Nidditch, ed. Oxford: Oxford University Press.
Needham, A., and R. Baillargeon (1993). Intuitions about support in 4.5-Month-Old infants. *Cognition*, 47, 121-148.
Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2003). Verifying properties from different modalities for concepts produces switching costs. *Psychological Science*, *14,* 119-124.
Povinelli, D. J. & Eddy, T. J. (1996) What young chimpanzees know about seeing. Monographs of the Society for Research on Child Development, No. 247, Vol. 61.
Prinz, J. J. (2002). *Furnishing the mind: Concepts and their perceptual basis*. Cambridge, MA: MIT Press.
Prinz, J. J. (2004). *Gut reactions: A perceptual theory of emotion*. New York, NY: Oxford University Press.
Prinz, J. J. (forthcoming). *The emotional construction of morals*. Oxford: Oxford University Press.
Prinz, J. J. and Clark, A. (2004). Putting Concepts to Work: Some Thoughts for the 21st Century. *Mind And Language*, 19, 57-69.
Rozin, P., Lowery, L., Imada, S., & Haidt, J. (1999). The CAD triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). Journal of Personality & Social Psychology, 76, 574-586.
Scholl, B.J., & Leslie, A.M. (1999). Explaining the infant's object concept: Beyond the perception/cognition dichotomy. In (E. Lepore & Z. Pylyshyn Eds.),, *What is cognitive science?* (pp. 26–73). Oxford: Blackwell.
Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990) Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, 13, 33-49.
Spelke, E. S. (1994). Initial knowledge: Six suggestions. *Cognition, 50*, 431-445.
Vinden, P.G. (1999). Children's understanding of mind and emotion: A multi-culture study.