I, Matthew Van Cleave,
hereby submit this work as part of the requirements for the degree of:

Doctor of Philosophy

in:

Philosophy

It is entitled:

The Role of Language in Cognitive Development

This work and its defense approved by:

Chair:  Christopher Gauker, Ph. D.
        Robert Richardson, Ph. D.
        Thomas Polger, Ph. D.
The Role of Language in Cognitive Development

A dissertation submitted to the
Graduate School
of the University of Cincinnati
In partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
In the Department of Philosophy
of the College of Arts and Sciences

By

Matthew Van Cleave
B.A., Psychology
Huntington University, 1999
M.A., Philosophy
University of Toledo, 2002

Committee Chair: Dr. Christopher Gauker
This dissertation concerns the role of language in cognitive development. I consider two kinds of continuity and discontinuity in cognitive development:

- **Continuity**
  - **Continuity\(_C\) thesis**: conceptual representation is built out of nonconceptual representation.
  - **Discontinuity\(_C\) thesis**: conceptual representation cannot be built out of nonconceptual representation.

- **Discontinuity**
  - **Discontinuity\(_L\) thesis**: acquiring a language makes possible new kinds of representation.
  - **Continuity\(_L\) thesis**: acquiring a language does not make possible new kinds of representation.

I argue for discontinuity\(_C\) and suggest a way to vindicate discontinuity\(_L\). I engage a number of different theorists in the course of this dissertation who likewise attempt to vindicate discontinuity\(_L\), but I argue that they are unsuccessful in doing so and that they ultimately embrace the continuity\(_L\) thesis. Although I do not defend it in this dissertation, I suggest that discontinuity\(_C\) and discontinuity\(_L\) may be the very same discontinuity.
TABLE OF CONTENTS

INTRODUCTION: DISCONTINUITY IN COGNITIVE DEVELOPMENT 1

CHAPTER 1: NONCONCEPTUAL THOUGHT, CONCEPTUAL THOUGHT AND ABSTRACTION 5
   The representation vs. the represented 5
   Judgment and Categorization 7
      Judgments and error 14
      Use of concepts in a range of different judgments 17
      Judgments and compositionality 19
   The problem of abstraction 20
      Comparative account of abstraction 26
   Nonconceptual thought 37

CHAPTER 2: EMPIRICIST CONCEPTS 40
   Infant categorization in Mandler’s work 43
   The *explanans* and the problem of abstraction 46
   Mandler’s account of abstraction 48
   Interlude: Is conceptual representation processed implicitly? 57
   Barsalou’s account of abstraction 63
   An *explanans sans* abstraction 73

CHAPTER 3: THE OBJECT CONCEPT 80
   Object individuation in infancy 81
      Experiment 1 83
      Experiment 2 84
   Is the object concept a concept? 91
      Carey and Xu’s argument 92
      The object concept measured against some characteristics of conceptual thought 99
   How does language affect individuation? 107

CHAPTER 4: UNDERSTANDING BELIEFS 113
   The vicissitudes of Jill de Villiers’s linguistic complements account 115
   An alternative account: The conversational account 123
   Objections to the conversational account 128
   Other supporting data for the conversational account 138
   Conclusion 140

CHAPTER 5: PHILOSOPHERS ON LANGUAGE AND THOUGHT 142
   Andy Clark: Cognitive offloading 142
      The role of language in cognition 149
INTRODUCTION: DISCONTINUITY IN COGNITIVE DEVELOPMENT

This dissertation concerns the role of language in cognitive development. I consider two kinds of continuity and discontinuity in cognitive development:

**Continuity**
- **Continuity thesis**: conceptual representation is built out of nonconceptual representation.
- **Discontinuity thesis**: conceptual representation cannot be built out of nonconceptual representation.

**Discontinuity**
- **Discontinuity thesis**: acquiring a language makes possible new kinds of representation.
- **Continuity thesis**: acquiring a language does not make possible new kinds of representation.

I argue for discontinuity and suggest a way to vindicate discontinuity. I engage a number of different theorists in the course of this dissertation who likewise attempt to vindicate discontinuity, but I argue that they are unsuccessful in doing so and that they ultimately embrace the continuity thesis. Although I do not defend it in this dissertation, I suggest that discontinuity and discontinuity may be the very same discontinuity.

In chapter one I develop and defend a distinction between nonconceptual and conceptual thought. I suggest that what characterizes nonconceptual thought is the ability to categorize whereas what characterizes conceptual thought is the ability to make judgments, which I characterize as the ability to think of a particular as belonging to a kind. I present four differences between categorization and judgment which suggest that they may be distinct abilities. In the second half of chapter one I attempt to defend discontinuity by arguing against continuity. If we are supposing that concepts are built out of nonconceptual representations, then an account must be given of how this happens. The problem I raise for continuity is the problem of abstraction: Why is it that I form concept X from my experience rather than some other concept Y? I introduce two
different kinds of account of abstraction, the comparative and noncomparative accounts, and suggest that the comparative account holds the most promise for an account of how we form concepts since it allows for a temporally extended process. The comparative account claims that concepts are created from stored representations of objects (what I call “vector fingerprints”) and that these representations are not conceptual. Similar vector fingerprints cluster together in a vector space, creating the concept. However, I argue that the comparative account faces a dilemma: if the dimensions that characterize the vector space are broad ones such as color, shape, texture, etc., then we cannot get the clusters to reflect the concepts we actually have. But if the dimensions that characterize our vector space are narrow ones then we run the risk of already presupposing concepts which are themselves at least as problematic as the ones we are trying to build.

In chapter two I examine two recent concept empiricist accounts of concept formation: Jean Mandler’s “image-schemas” account and Larry Barsalou’s “perceptual symbols” account. I show that neither account is able to avoid the problem of abstraction. In addition, I argue against Mandler that we need not attribute concepts to prelinguistic infants in order to understand their behaviors in certain experimental paradigms. The significance of this argument, here and elsewhere in the dissertation, is that it leaves open the possibility that conceptual thought may depend on language. If we are not required to attribute concepts to prelinguistic children (or nonlinguistic creatures), then it may be the case that discontinuity_C maps onto discontinuity_L. But if we are required to attribute concepts to prelinguistic children, then discontinuity_C could not map onto discontinuity_L.
In chapter three I argue against Fei Xu and others who claim that children have an early developing object concept (more precisely, a concept of what are called “Spelke-objects”). Rather, I suggest that the behaviors in various experimental paradigms can be understood by attributing only nonconceptual representations, including a kind of imagistic subitization ability. The significance of this is, again, that we are not required to explain the behaviors of prelinguistic children by attributing concepts, and that this leaves open the possibility that conceptual thought may depend on language. In this chapter I also examine Xu’s “essence placeholder” account of how language creates kind concepts and find it lacking in several ways. In particular, Xu’s account leaves no significant work to language and so is best seen as embracing continuityL. In addition, it seems she must embrace continuityC since all the real work of forming kind concepts is done by language-independent mental mechanisms. Thus Xu’s account leaves no room for language to make any kind of significant difference to thought.

In chapter four I examine Jill de Villiers’s account of how language might enable one to solve the false belief task. Children under three years of age are notoriously inept at attributing false beliefs to others (as well as themselves). De Villiers suggests that language has an important role to play in children’s eventual success (typically by five years old) at the false belief task. She offers several accounts according to which mastering a particular syntactic category of natural language sentences is the key element for being able to represent that someone falsely believes something. However, I argue that none of her theories is successful. I offer an alternative account of how language may play a role in children’s coming to understand belief. On the account I offer, it is the interpersonal function of language that is key to understanding this particular aspect of
cognitive development. The significance of this argument is that it suggests a way to vindicate discontinuity$_L$: the particular importance of language is its interpersonal function. In chapter six I suggest we look to this kind of account if we wish to vindicate discontinuity$_L$.

In chapter five I engage four philosophers who attempt to provide a role for language in cognitive development: Andy Clark, Daniel Dennett, Jose Bermudez, and Peter Carruthers. I argue that none of their attempts to provide a significant role for language is successful. In particular, each theorist fails to vindicate discontinuity$_L$. In the last chapter I offer an analysis of a common assumption that each of these accounts share. It may be that this assumption is what is prohibiting these theorists from vindicating discontinuity$_L$. In addition, I argue that the evidence that Bermudez offers for nonlinguistic creatures having concepts could be given an alternative explanation in terms of nonconceptual thought.

Finally, in chapter six I argue that none of the accounts of the role of language in cognitive development surveyed in chapters three, four and five is successfully. Furthermore, I argue that they all attempt to see the significance of language for cognition in terms of its intrapersonal function rather than an interpersonal function. I suggest that in light of the failure of these accounts to vindicate discontinuity$_L$, perhaps we should look to the interpersonal function of language in attempting to vindicate discontinuity$_L$. 
In this chapter I will develop some philosophical distinctions regarding the nature of thinking and cognitive development, which I will apply, in subsequent chapters, to specific proposals from psychologists and philosophers. The key distinction I wish to defend is a distinction between conceptual and nonconceptual thought. I will distinguish conceptual thought as the ability to make judgments and will describe several ways in which the ability to make judgments differs from the ability to categorize. Any account which attempts to explain how we acquire the ability for conceptual thought must face the problem of abstraction, which is the problem of explain how, from experience with particulars, we can acquire representations of kinds. I will lay out this problem in the second part of the chapter.

The representation vs. the represented

One possible confusion in a discussion about representation is an equivocation with words like “information,” “aspect,” and “feature.” The potential equivocation is to use these terms as referring to both the thing represented and the representation of that thing. The mistake is to assume that one can unproblematically talk about representations of a thing just because one can unproblematically talk about the thing itself. For example, I can talk about beaks, which are clearly a property ascribable to birds, and I can talk about representations of beaks, which may be a part of my concept BIRD. An equivocation would be made, however, if I assumed that my license to talk about the property of being
a beak licensed my talk about representations of beaks. Of course, in giving an account of representation, we can assume the existence of things in the world like the beaks, but we can’t assume that just because there are beaks in the world we must also have a representation BEAK. Rather, the representation, BEAK, is precisely what needs to be explained on any account of representation. In general, one is licensed (in psychology and cognitive science, even if not metaphysics) to assume the existence of certain properties in the world such as wings and beaks and colors and shapes, but one is not thereby licensed to assume the representations of those properties. One of the biggest problems in psychology and the philosophy of psychology is to explain how one acquires the ability to think thoughts about the world. But the equivocation has the effect of just assuming an answer to that problem since (according to a dominant view) thoughts are made up of representations of the world.

A common candidate for this equivocation is a word like “information.” For example, the psychologist Jean Mandler remarks:

Procedural knowledge, both perceptual and motor, is inaccessible to consciousness. It is unselective, taking in all the encodable information that is presented to the input system. This information is processed in parallel…(2004, p. 54, my emphasis).

On the one hand Mandler talks about information as impinging on the input system, which means it must be the things in the world, while in the very next sentence she talks about information as something that is processed, which means it must be something in the mind. In order to be clear about what one is talking about, it is common to adopt
some convention to distinguish the representation of the thing from the thing itself.¹  I will here adopt the convention of writing in capital letters when I’m talking about the representation, quotations when mentioning words and just plain lower case letters when talking about properties in the world. So, for example, birds have beaks and my concept BIRD may be composed of BEAK and WINGS.

**Judgment and categorization**

Conceptual thought is thinking that uses concepts. Both “thinking” and “concept” are terms whose definitions are highly disputed and in this section I will defend a particular view of how they should be understood. In later chapters I will set this characterization to work on various claims made by both psychologists and philosophers about what conceptual thought is and how one might acquire it. My claim is that conceptual thoughts form a subset of things we classify as thoughts and that what distinguishes conceptual thought from other kinds of thought (nonconceptual thought) is the ability to make judgments in the subject-predicate form, X IS Y, where, in the most basic case, X is a particular-specifying term and Y specifies a kind to which that particular belongs. For example, THAT [indexical] IS A DOG.²

How are we to define “concept?” Many people could be made to agree that concepts are mental representations which capture some understanding of a kind of thing. That is, it stands for more than one object. For example, having the concept DOG means

---

¹ Mandler adopt this same convention herself to refer to her image schemas, but at other times she still sometimes slips from talking about the thing in the world to the representations without being explicit that she’s doing so.

² Of course, the subject can also be a kind term in addition to an individual specifying term, as in ALL DOGS ARE MAMMALS. But the case in which the subject is a particular specifying terms seems to be more basic. C.f. Strawson, 1974.
that I can represent not just this dog but dogs in general. But what does it mean to represent “dogs in general?” I claim that being able to represent dogs in general means being able to think of a particular (e.g. this object) as belonging to a kind (e.g. DOG). Thus having the concept DOG entails at the very least being able to think a predicative thought like X IS A DOG. Such predicative thoughts I will call judgments. Thus, my claim is that if you have a concept then you can make judgments using that concept.

Much dispute, of course, arises over what we are willing to say counts as representing a particular as belonging to a kind. The clearest evidence would be that one can express the concept DOG by using “dog” in various different kinds of sentences like “that is a dog,” “dogs are animals,” “I was once bitten by a dog,” etc. Perhaps there are other nonlinguistic behaviors which are sufficient for us to count a creature as representing such things, even if they can’t say it in words. But which behaviors would be able to tell us? Does the frog represent the fly as food? Does the monkey represent the lion as a lion or a certain kind of predator? Does the connectionist network represent the rock as a certain kind of object? Perhaps the frog responds in a particular way (eating behavior) to all flies and perhaps the monkey responds in a particular way (e.g. manifests predator avoidance behavior) to all lions. In that case, we could say that the animal is representing all the instances of flies (or lions) as being in some sense the same, since they respond the same way to all the particular instances. Of course, there’s the problem of what counts as the same type of behavior, but that is a complication I will disregard. I assume that we are pretty good at typing kinds of behavior in pretty broad respects, such as predator avoidance behavior.
The problem with all of these cases, however, is that although the animal or artificial system may respond in ways which make us want to say they are representing a kind of thing, the animal itself need not represent this fact in the form of a judgment such as “x is a rock.” And although at this point I’m willing to grant that judgments could be made nonlinguistically (perhaps using images—something which I will dispute in chapter 2 when considering Barsalou’s theory), I think it’s pretty clear that having a particular kind of response to a particular kind of thing is not sufficient for a system to be said to represent a particular as belonging to a kind. This is clear from examples like Churchland’s “rock-mine” detector, which differentially responds to rocks and mines, but which clearly doesn’t think and doesn’t make judgments.

The rock-mine detector is an artificial connectionist network which is trained to classify various instances of a sonar echo profile as either a rock or a mine (think submarines). Here’s how it works. The input level of the network (which is functionally equivalent to sense organs) reads the frequency pattern at thirteen different points; each frequency in the whole pattern will have a certain level or power from 0 to 1. For example, one frequency, say 440, will have a power of .14. The input layer is connected to a hidden layer (seven nodes) and then to the output layer which consists of only two nodes, one representing mines and one representing rocks. Each layer has connections to every node in the next layer, and each connection has different weights (i.e. signal strength). The interesting thing about networks like this is that they can learn to discriminate a mine echo profile from a rock echo profile (i.e. it produces an output vector <1,0> when a mine is present and <0,1> when a rock is present). It learns to do

---

3 This example is taken from “On the Nature of Theories” (Churchland, 1989).
this by being fed particular echo profiles and then being taught to respond correctly. Teaching it to respond correctly means adjusting particular weights and seeing if this adjustment brings the output closer or further away from the desired output (i.e. to $<1,0>$ if the input was a mine).\footnote{Actually, none of the weights are adjusted until appropriate error has been assigned to each of the nodes having weighted inputs, since in order to calculate the error of the hidden units, one must refer back to the combined and weighted error of the output units (relative to the input in question). But if one had already made adjustments to the weights leading into the output layer, then one could no longer refer to the error in outputs. See Plunkett and Elman, 1997, p. 13.} This plodding method is called learning by backpropagation of error. An important point about backpropagation is that the delta rule, which is an algorithm to calculate how the error (the difference between the desired output and the actual output) is affected by the adjustments of the weights, can be carried out by a computer. The delta rule makes training the network a tractable task, since it would take a human years to train a network manually by backpropagation. Once the network is trained it can correctly identify new samples as either a rock or mine. It can do this because the training has in effect partitioned the hidden unit vector space into a formation such that mines and rocks have a prototypical “fingerprint.” Thus, the ROCK fingerprint is a range of different hidden unit vectors which set off the ROCK output and the MINE fingerprint is a range of hidden unit vectors which set off the MINE output.

The input layers on any connectionist network will correspond to various dimensions to which the “sense organs” are sensitive. In the case of the rock-mine detector there is only one dimension, sound frequency, which is divided up into different units, which detect specific sound frequencies. A different connectionist network might have an input layer that is sensitive to color, in which case the input units would be divided up to detect different light frequencies. The input layer, in both humans and in
connectionist networks, is hardwired. The hidden layer, in contrast is where the “learning” takes place and where the representations ROCK and MINE reside. But it is important to be clear about the kind of learning that is going on in a three-layer feedforward network like the rock-mine detector. Most importantly, the account by which the network learns to classify something as a rock or mine depends on there being a representation of a desired output. But the system itself isn’t representing the desired output, since if it was it would already have the concept it was trying to learn. Rather, the programmer (via the backpropagation algorithm which trains the network) has specified for each input what the desired output is for each instance. So although there is a sense in which the network has learned to classify rocks and mines, it is a kind of learning that depends heavily on a teacher, who specifies for every encountered instance, the kind to which that instance belongs. It is precisely the extent of this kind of dependence on a teacher which makes backpropagation an unrealistic method of learning for humans (or other animals), who don’t get this kind of feedback from the environment.

I think that the rock-mine detector doesn’t count as a conceptual thinker even though it is a categorizer. Its categorization abilities put it in a class with other categorizers, such as pigeons, which can learn to respond differentially to different kinds of objects. But although a connectionist network may even be a plausible account of the categorizer’s innards, it doesn’t seem to have the resources to account for conceptual thinking. The main problem is just that the connectionist network doesn’t have the kind of structure to be able to represent a subject-predicate judgment. For each instance it is presented, it correctly puts in into a category, but it cannot represent the particular as belonging to the kind.
One manifestation of this problem is that it cannot represent the particular without at the same time representing it as a kind. So, for example, the rock-mine detector would, speaking loosely, locate the particular input vector in the ROCK hidden unit space, but it couldn’t separately refer to that same individual because its way of referring to the individual is exhausted by its method of categorizing it. For example, we could imagine a more complex network which had more than one kind of “sense organ.” Imagine that we added a sense organ which was (somehow) able to encode a texture vector for the very same objects for which the network encodes “echo profiles.” Now the question is: how could it say of an object which it had categorized as a ROCK, that that same object was also ROUGH? The problem is that if the way the network is supposed to refer to the particular is by its input vector, then how are the two different input vectors (for the texture and the sound) supposed to be about the same object? The network could not classify the object it had classified as a ROCK as also ROUGH. It could not apply two different concepts to the same individual. And the reason seems to be that the network can only refer to the particular within the limited context of the specific input vector, which already classifies it.

One might object that, nevertheless, there is indeed enough structure since the input vector could count as the representation of the particular and the partitioned hidden unit space could count as the representation of the kind. Even if there is a problem for more complex networks that contain more than one input dimension (viz., that it can’t apply two different concepts to the same individual), simpler networks would have the resources to represent particulars (as inputs) and kinds (as portions of partitioned hidden

---

5 It seems to me that the problem I am raising here has some similarities with the so-called binding problem in vision.
unit space). However, a connectionist network does not itself have a representation of its own hidden unit space, which is shown by the fact that it can’t generalize over the “concepts” (partitioned hidden unit space) it has (c.f. Clark and Karmiloff-Smith, 1993). Thus, although the rock-mine network correctly identifies any particular input vector as either a ROCK or a MINE, it could not represent, e.g., what was characteristic of all rocks. Suppose, for example, that all mines were unique in that they all registered a spike at a frequency of 440. What the network could not do is represent this fact, and the reason it couldn’t do it is that representing this fact would require representing its own partitioned hidden unit space (its representation MINE). That a first order connectionist network cannot do this is shown by the fact that the “knowledge” embedded in a network is domain or task-specific and cannot be transferred to a new task or domain (Clark and Karmiloff-Smith, 1993).

The moral of the story is that having the ability to differentially respond to particular kinds of stimuli in one’s environment, such as the connectionist network has, is different from being able to think a thought. Someone who claimed that the rock-mine detector does have the concepts ROCK and MINE would just be applying “concept” more liberally than I am, since he would apply it to things that are neither making judgments nor thinking. But the notion of concept I’m interested in is that which relates to thinking thoughts and so there will have to be something more to the story than merely an ability to respond to a particular kind of stimuli. Of course, something might have the ability to respond differentially to kinds of things as well as the ability to think thoughts
about the things in question, but then it’s the thinking of thoughts that would potentially count as the exercising of concepts, not merely the differential response.⁶

There are two abilities which I claim are separable: the ability to categorize and the ability to form subject-predicate thoughts (judgments). Sometimes concepts are identified with an ability to categorize, but although there may be mental representations which account for one’s ability to categorize, this categorization ability is not sufficient for thinking subject-predicate thoughts. One reason for this has already been given: there are some things which clearly have the ability to categorize but which lack the ability to make judgments.

**Judgments and error**

Another thing that distinguishes categorization from judgment is that with judgment comes the possibility of making an error, something which is not within the purview of mere categorization abilities. One of the key marks of a judgment such as X IS A VEGETABLE is that we can be mistaken in making it. If I judge that a tomato is a vegetable then, in fact, my judgment is false and I can take steps to correct this judgment based on the feedback I get. But suppose I’m sorting the tomatoes in terms of which ones I would like to eat and which ones I wouldn’t (or ordering them in terms of their color, or in terms of their softness, etc.). In such cases, what would it mean for me to make an error? If something looks like something I’d like to eat, then it is. Likewise, if object x

---

⁶ One might think that it’s the complexity and subtlety of the responses and the ability to create new ways of responding that warrants the inference that judgments are being made. For example, if an animal were to demonstrate an ability to make a novel inference, we might want to say they are making some kind of judgment.
looks more similar to object y than to object z with respect to color, then how could I be mistaken in ordering them as I do?

Someone might object here that there is a sense in which I could make an error in sorting. For example, suppose I put a rotten tomato, which will make me sick, into the “eat” pile. In that case, I will have made an error since if I knew it would make me sick then I wouldn’t have put it into the pile. Furthermore, we can suppose that the rotten tomato had certain distinguishing marks (which initially led me to put it into the good pile) and that after eating this tomato and becoming ill, I would no longer sort tomatoes with those distinguishing marks into the “eat” pile in the future. In this case, it might be argued, it seems that not only would we be making an error (objectively, we might say) but we could also become aware of the fact that we made an error because of the effects of eating the rotten tomato. Here I would claim, however, that recognizing the effects of some action is different than recognizing that an initial judgment was mistaken. But it is the latter, the ability to recognize that one’s initial judgment was mistaken, which is characteristic of the kind of error which becomes possible with conceptual thought. This ability to recognize error requires the ability to connect up in one’s own mind the original action or thought with a subsequent thought in such a way that the subsequent thought acknowledges that the prior behavior or thought was mistaken.

So, in the case I am imagining, would we be able to say that the subsequent avoidance behavior of tomatoes that have certain distinguishing marks was a recognition

---

7 Davidson has made a similar point: conceptualization, unlike mere discrimination, requires the ability to recognize a mistake as a mistake (1997, p. 25).
8 In chapter 4 I will discuss a version of the false belief task, in which three-year-old children are unable to recall past beliefs which are now known to be false. My present criterion would imply that these children aren’t making judgments since they do not seem to have the ability to recognize that a past judgment was mistaken.
that one had made a sorting error initially? Of course, if we also had the ability to make judgments such as X IS A GOOD TOMATO TO EAT, then we could (after eating it and getting sick) recognize that the initial judgment was mistaken. But in order for the sorting behavior to have been mistaken, we would need to find a way to connect the subsequent avoidance behavior to that initial sorting behavior in such a way that it is recognized that that very sorting behavior was mistaken. But it seems more likely that the avoidance behavior be would tied to the particular perceptual marks of the tomato rather than to the initial sorting behavior itself. But in that case the avoidance behavior does not connect up with the initial sorting behavior and so wouldn’t be a recognition that the initial sorting was mistaken.

That categorization doesn’t have room for any account of error is even clearer in the case of sorting tomatoes according to their color or softness. It seems clear that in these kinds of cases, whatever seems right to the categorizer will be right. How could one be wrong here? If we imagine that the individual has a kind of vector representation space (or a similarity space—see chapter two) then depending on which aspect is most salient in the sorting, the tomatoes will be arranged according to the metric of the vector representation space (i.e. according to their particular vector values). If one’s vector representation space is responsible for the sorting/ordering of objects then in what sense could one’s ordering of the objects fail this very same sorting criteria? It seems there must me some independent criteria (apart from the vector space), which the sorting could fail to exemplify, in order for there to be a genuine possibility of error. But what criteria

---

9 One thing that is needed here is an account of how salience of different dimensions can change in different contexts, such that in some cases certain dimensions are weighted more heavily than others. I don’t have such an account, but I briefly return to this question in chapter 6 when considering an idea from Smith and Heise 1992.
If correctness is merely what seems correct to the individual, then how could what seems correct to the individual ever be mistaken (in which case, how is it possible to speak of correctness)? The problem is that the individual’s sorting behavior will de facto seem correct to that individual (that is why they sorted it that way in the first place), but in order for there to be a recognition of error, there must be some independent fact (other than the seeming correct) which can act as a measuring rod against what seems correct to the individual.

*Use of concepts in a range of different judgments*

Another reason categorization and judgment are different abilities is that there are different requirements for what counts as exercising them. The ability to judge depends on being able to exercise the concept involved in the judgment in a number of different circumstances, including circumstances in which the property judged is not sensibly present. The ability to categorize requires only that one have a relatively stable disposition to discriminate one property from a range of others. Consider, for example, one way in which it might be suggested we learn the concept BANANA. Upon seeing many bananas and hearing the word “banana” one learns the concept BANANA. Does one thereby acquire the ability to judge “x is a banana?” One might wonder how to tell the two abilities (categorization and judgment) apart in this case.

---

10 Adrian Cussins 2003 has suggested an analogue of the truth norm for nonconceptual thought—something like action guidedness…

There is a Grey parrot, Alex, that can answer questions about color and shape of objects when asked.\(^\text{12}\) For example, when Alex is asked “how many blue blocks?” he answers correctly. However, one must be careful to distinguish whether Alex is actually making judgments or has merely learned, via operant conditioning, to connect certain sounds with certain sights without thereby understanding what the words mean. (One might wonder if Alex the parrot is in a kind of Chinese room (Searle 1980) situation where he learns to connect one string of meaningless symbols to another string of meaningless symbols.) Of course, Alex has to be able to connect a certain visual stimulus with a certain string of sounds, and that is a feat, but it’s not a feat that requires the ability to represent a particular as belonging to a kind. If someone truly had a concept then they would be able to use it in novel ways that weren’t restricted to being triggered by the same environmental stimulus every time. To have a concept is to be able to use it in some situations where nothing corresponding to that concept is sensibly present (c.f. Geach 1957, p. 35). If someone could only use the concept RED in a highly restricted and artificial setting in which there we always red things present, our confidence that they understood how to use the concept RED would be undermined. Likewise, if Alex could only utter the word “rock” when shown a rock, then this should undermine our confidence in attributing to him the concept ROCK.

But Alex can also ask for things which aren’t sensibly present, such as a banana. And it seems that he has a banana in mind in light of the fact that he will not accept other kinds of food such as nuts (which, if given, he will throw back a the experimenter). Thus, Alex’s use of “banana” seems to be a command to the effect of “give me a

\(^{12}\) These studies have been carried out over a number of years by Irene Pepperberg. Summaries of this work can be found in Pepperberg (1999, 2002).
banana.” But we need not think that Alex thereby has concepts of ME and GIVE in order to understand what is going on here. Rather, Alex may just feel hungry and recall an image of the experimenter giving him a banana and connect the image of a banana with the sound “banana.” But in this case he’s using the word “banana” as intimately connected with the desire to eat a banana, and thus he needn’t have the ability to use “banana” in other contexts such as saying which fruit is the largest, or which fruit is yellow. But if one’s use of a term cannot be separated from a particular and limited context, then the worry is that what’s going on does not involve the grasp of a concept.

One of the characteristics of having a concept is that one can employ it in many different kinds of contexts. So, if I have the concept BANANA then not only can I ask for a banana, I can also tell you that the banana is the yellow one or that the banana is the longest one. Unless Alex also had abilities such as these one should be skeptical of saying that Alex is judging “x is a banana.”

Nevertheless, Alex does have an interesting ability in asking for the banana since he is essentially using the word to make something happen, to do something like make a command. In this sense he may have the beginning of a use of the word “banana” and perhaps the beginnings of a concept. I say “beginning” because there’s as of yet no account of how Alex could misrepresent a banana, and misrepresentation is essential to be able to say that judgments are being made.

Judgments and compositionality, productivity and systematicity

Another difference between acts of judgment and acts of categorization is the often noted requirement that judgments exhibit productivity and systematicity, and that, therefore, the
concepts which constitute the judgment exhibit compositionality. The basic idea is that it is definitive of concept possession that one be able to recombine concepts in new ways. Suppose I have the concepts BALL, BOX, RED, and BLUE. If I can think the thought that the ball is red and the box is blue then I can also think the thought that the box is red and the ball is blue, even if I’ve never encountered a blue ball before. If I have a concept, I need not receive explicit training in how to apply that concept in every possible way. That children can use language productively in this way was Chomsky’s argument against Skinner’s account that language learning can be accounted for in terms of operant conditioning (Chomsky 1959). The idea is that it is precisely because thoughts are compositional that thinking is productive, i.e. that we can contemplate an infinite number of different and new thoughts. However, something might have the ability to categorize (respond differentially to) a blue box and a red ball without thereby having the ability to categorize (respond differentially to) a red box or a blue ball. Of course, we could possibly train the system to recognize a red box or a blue ball, but then this would have to be something extra we trained it to do. This, in effect, is Fodor’s argument against distributed representations in connectionist systems being able to serve as concepts.13 My point here is more general: the ability to categorize does not require compositionality and productivity, but the ability to use concepts as constituents in judgments does require that they can be used productively.

**The problem of abstraction**

If conceptual thought is most fundamentally being able to think of a particular as belonging to a kind, then the pressing question that needs to be answered is: how do we come to represent kinds? Any account according to which concepts are mental representations must explain how a mental representation can stand for a set of objects corresponding to a kind. For example, if I have a concept DOG then (on the mental representation view of concepts) I must have a mental structure which stands for all and only dogs. It is precisely because the mental structure represents all and only dogs that I can make judgments about particular dogs. When I judge that “that is a dog” then I am asserting something that can be either true or false. Its truth or falsity depends on how the world is, and a common way to understand truth and falsity is in terms of the reference relation. We say that the concept “dog” refers to dogs and that the truth of the judgment “that is a dog” depends on whether the object to which you refer is in fact a dog. The ability of a kind to represent all and only members of that kind is, then, on this view, the feature which allows for the characteristic feature of judgment, truth and falsity.

The problem of abstraction is the problem of explaining how kind representations are created from experience with particular objects. Of course, one could just claim that all concepts are innate, but there is something so counterintuitive about such a rampant nativism that most philosophers and psychologists think that most concepts are learned. Thus one constraint on any account of abstraction is just that it must avoid appealing to innate dispositions in accounting for most of the concepts we have. Some innate dispositions may be (and perhaps must be) admitted, but not to the point of positing one innate disposition which generates each concept we acquire.
Another key constraint on any account of abstraction is explaining the generation of concepts in a way that makes it clear how we can *share* concepts. One of the characteristics of concepts is that they enable us to communicate. If we are able to communicate with other speakers of our language then it can’t be that we have radically differing concepts. One of the challenges facing the proponent of the mental representation theory of concepts is to explain how we would each individually come up with the same concepts. Of course, one might very well explain how we all come up with the same concepts if one could appeal to innate dispositions to form certain concepts (and not others), but this would violate the first constraint.

It seems there are two different kinds of account of abstraction one might give: a comparative account or a noncomparative account. On the one hand, one might say that abstraction operates by comparing various representations of particulars and seeing what they have in common, that common property (or properties) being what is abstracted. On the other hand, one might say that abstraction operates by creating a general representation from just a single perception of an object (without comparing that perception to other perceptions in order to see what that collection has in common). It seems that the comparative account initially holds more promise of accounting for how concepts are learned, since it (unlike the noncomparative account) countenances a temporally extended process which would seem to be required for any kind of learning to take place. The noncomparative view, according to which concepts are acquired in “one-shot” learning, seems to depend on there being some recipe for how to construct the concept from the percept already contained somewhere within the person’s mind. But to say that the recipe for the concept is there before one even has any experience with
instances of that concept seems no different than claiming that the concept is innate. Even if there is a sense in which one doesn’t have a concept until one encounters an instance of the concept, at which time the concept is “activated,” it is still the case that the concept isn’t learned at all, rather, it is already there waiting to be activated (c.f. Fodor 1998).

Nevertheless, both accounts face the same problem initially: why is it that concept X is abstracted rather than some other concept Y? On the noncomparative view, there are many different things I could abstract upon seeing, e.g. a dog. I could abstract DOG HEAD or ANIMAL WITH EYES or SPOTTED DOGS or FOUR-LEGGED ANIMAL. What makes it the case that I abstract the concept DOG? Likewise, on the comparative view, from the objects I’ve perceived there are many classes that could be formed whose members have some feature in common. For example, probably all the dogs I’ve seen belong to the class of animals smaller than a cow. Most of them belong to the class of furry household pets. Most belong to the class of animals that walk on all fours. Some belong to the class of small dogs and large cats. Some belong to the class of dogs in Cincinnati. The question is what determines which concepts I am to abstract from the objects I have encountered, given that there are multiple concepts that could be abstracted? For example, since the particular dogs I have experienced belong to multiple different classes, how do I come to abstract the concept which picks out the extension of DOG? The problem is not only that there are many different properties to abstract from the class of dogs, but also that we have an initial problem of figuring out which class of objects we are supposed to abstract a concept from. It’s not the case, for example, that I would already to know just to focus on dogs in order to abstract some common property.
Rather, starting from scratch, I would have one large class of objects (including cats, stuffed animals, ponies, etc.) for my abstractive mechanism to work from.

In order to be able to fulfill all the roles a concept is supposed to play in communication and thought, abstraction needs to be a process which creates a concept whose extension contains all and only members of the kind. The problem is that it’s not as if there were only one extension to which the exemplars I’ve encountered belong. Thus one must explain how we abstract the concept DOG from our experience with dogs and cats and other things as well. Any account which attempts to show how concepts are created from percepts must offer an account of how the particular concept they are trying to explain is created from the “blooming, buzzing confusion” of experience.

Thus stated, the problem of abstraction is akin to a curve fitting or underdetermination problem: there are many different theories which are consistent with the data. Just as one could draw indefinitely many different curves through a set of data points, one could abstract indefinitely many different concepts from one’s experience with various different particulars. Unfortunately, in the case of abstraction, unlike in the case of scientific theories, one cannot appeal to some notion of simplicity in order to “weed out” the “weird” concepts and leave us with only the acceptable ones, i.e. the ones we actually have, like DOG. Scientific theories appeal to simplicity, predictive utility, elegance and other pragmatic and aesthetic considerations. But the mechanisms purportedly responsible for abstraction would not themselves narrow down the range of alternative conceptions based on such considerations. Those mental mechanisms (if there are such) do not have ideas about what is simple or elegant on the basis of which they

14 I owe this analogy to Dan Weiskopf.
would choose one conception over another. One might think, though, that inasmuch as one conception would be more advantageous to hold than another—perhaps because of its predictive utility—then evolution would have designed our abstractive mechanism to throw out the options with, say, less predictive utility. For example, suppose a certain kind of prey formed concepts about one of its predators which was typically spotted, but not always so. And suppose that all the instances of this predator that the prey had observed were spotted. Under such conditions, perhaps the abstractive mechanism would develop an adaptation which would abstract a concept of the predator that wasn’t limited to it being spotted (i.e. a wider concept), since then, if it were to encounter a non-spotted version of the predator, it would recognize it as such and respond accordingly.

But this way of narrowing down the space of options faces a dilemma. For if the abstractive mechanism is supposed to only rule out some alternative concepts but not all, then there will still be a problem of selecting between alternative concepts. On the other hand, if the abstractive mechanism is supposed to completely narrow down the space of options to a unique concept, then the concept is innate. So attempting to arrive at a unique concept by appealing to adaptations of the abstractive mechanism either ends up making concepts innate or leaves us with the same underdetermination problem again.\(^{15}\)

How, then are we to narrow down the options to arrive at a unique concept without thereby claiming that the concept is unique? It seems to me that the noncomparative account of abstraction really has no chance of explaining how concepts are learned. The reason, given above, is that there’s no time and hence no place for any

\(^{15}\) There is a further problem if this account of supposed to describe human concepts, since even if we did acquire some concepts innately in this way, it is implausible to think that all our concepts are adaptations.
kind of process of narrowing down or winnowing away at the space of options. Although it may be that not every possibly way of grouping the instances is an ecologically valid way of grouping them (for example, if I don’t already have the concept CINCINNATI, then I wouldn’t be able to form the concept DOGS FROM CINCINNATI), there will still be many differing conceptions that fit the “data.” Thus, if there’s going to be an admissible account of concept learning, it must be from the comparative account.

**Comparative account of abstraction**

I think that a common intuition behind the comparative account of how kind representations are created is that one notices the similarities shared by, say, all and only dogs and not shared by any other thing. No two dogs share all the same properties but, it is claimed, all dogs share some properties, and it is these shared properties that constitute our concept DOG.\(^{16}\) Somehow we are able to disregard the inessential properties, such as the color of the dog, and pay attention to the essential properties—those properties that only dogs share—perhaps things such as tail wagging, barking, etc.

At this point, however, it is important to note two ways of going awry. First, one must be careful not to presuppose other concepts in the building of a particular concept. In this case, one cannot leave unexplained the constituents, BARKING and TAIL WAGGING, of the concept DOG, since the question of how we represent tail wagging and barking is just as crucial a question as how we represent dogs. The general point is

\(^{16}\) This account is supposed to be neutral with regard to the particular mechanisms by which concepts represent kinds. So for example, a concept might be a prototype or an exemplar. However, on any of these theories there is some unique group of properties in virtue of which something is a dog, and our concept DOG is somehow keyed into these properties.
that any purported account of concept acquisition must be careful not to posit concepts which are themselves left unexplained. Thus, he must not say that, for example, the concept BIRD is formed by noticing that birds have WINGS and BEAKS whereas nothing else does (as Eleanor Rosch does, see Rosch et al., 1975; Rosch, 1977). It may be that once we have the concept BIRD and BEAK and WING we define the concept BIRD in terms of WINGS and BEAK. But since WING and BEAK are themselves concepts, we would either have accept them as innate concepts or explain how we acquired them. But any attempt to explain how we acquired BEAK and WING along the same lines faces the looming regression of having to explain what all beaks or all wings have in common (i.e. one can’t very well explain wings in terms of FLYING, without having to explain how we get that concept, etc.). This point applies equally to the noncomparative view of abstraction. Second, one must not presuppose the very concept whose formation one is trying to explain. For example, if it is said that I abstract the concept DOG by noticing what all and only dogs have in common, then it seems I am presupposing I already have some way of telling what all and only dogs have in common, whereas this is what needs to be explained.

If we can’t build our concept DOG out of parts like TAIL WAGGING (since these themselves are concepts), perhaps we could build them out of something “simpler.” Here the suggestion is that we clearly have the ability to discriminate various different kinds of properties (even pigeons can be trained to respond to different color patches), so perhaps we can build up a concept out of simpler properties which we could plausibly discriminate without having to have a concept thereof. But what might these “simpler” things be? The foregoing criticism of the comparative account of abstraction was that if
the things compared are described in a way that already employ other concepts, then we are stuck with the problem of explaining how we got those concepts. Thus, it seems that if we are to avoid an infinite regress, we must at some point be able to describe the instances (which are being compared) in a way that doesn’t already presuppose other concepts. In chapter two I will consider Lawrence Barsalou’s account of how this might happen. For now, I will just present a general outline of how such an account might go and the problems it must face.

The basic idea is that we store copies of our sensory experiences of the instances we encounter. Perhaps we store these copies with all of their particular accidents or perhaps we store more schematized versions. Either way, no two instances will be exactly the same. There must also be some mechanism or gate for allowing things to be stored since presumably not everything that our senses pick up is thereby stored in long-term memory. A not uncommon idea is that conscious attention is such a gate, with those things being consciously attended to being stored in long term memory. The question now is how are we to describe these copies of the instances stored in long-term memory? The basic idea is that we store these instances as a kind of vector whose dimensions are certain ranges which our sensory systems themselves are sensitive to. For example, we can distinguish the color of a thing from its shape, and its shape from its texture, its texture from its sound. So, for example, we would store the instance of a dog as a kind of vector according to the different dimensions of the vector—its shape, its color, its sound, its texture, etc. One might wonder if these “dimensions” of the vector are themselves concepts, in which case the prior criticism would not have been avoided. The reason they are not concepts is that they are ranges which characterize all objects, not groupings of
objects. All objects have some color, some shape, some texture, but not all objects have wings, legs, tails, etc. Thus characterizing an instance by its vector does not in itself group an object at all, it just places it in a vector space.

So, then, how might there come to be groupings of objects? The idea is that the vectors corresponding to kinds will have similar “fingerprints,” and as a result of their similarity they will cluster together, and these clusters will stand for the kinds. For example, dog1 and dog2 will have similar vectors in virtue of their similar shape, texture, movement and sound patterns, and since those vectors are similar, they will cluster together in the vector space. This cluster, then, will be the concept DOG. This, then, is one way to flesh out a comparative account of abstraction. The two important parts to the account are: 1) the vector representations of particulars, the “description” of which doesn’t invoke other concepts and 2) the idea that these vectors cluster together and that these different clusters represent different kinds.\footnote{17}

This account faces a dilemma, the first horn of which is this: if the dimensions which characterize our vector space are broad ones like color, shape and texture, then we will not be able to form the categories we do in fact have. The problem is that the clusters formed will be both too wide and too narrow to correspond to the categories we actually do have; that is, the clusters will include things which don’t belong to the category. For example, perhaps some dogs have shape and texture vectors more similar to cats than dogs. In this case the CAT category would be too wide and the DOG category too narrow. One might think that dogs could be excluded from a cat category on other grounds, for example, that they bark. The response to this is that we’re dealing

\footnote{17} Examples of this kind of approach include Barsalou (1999, 2005), Churchland 1989, and Gardenfors 2000.
with a vector, not a definition in terms of necessary and sufficient conditions (since such a definition would seem to already presuppose concepts), and the vectors are supposed to cluster together because they have similar vector patterns. But it seems that something need not be highly similar on all dimensions if it is highly similar on other dimensions. So, for example, the fact that dogs emit certain noises that are different from the noises cats emit wouldn’t in itself exclude the dog if the dog were highly similar on other dimensions such as shape and texture.\textsuperscript{18}

As another case in point, consider the problem of accounting for the concept CHAIR in terms of the aforementioned dimensions of vector space, i.e. in terms of SHAPE, COLOR, TEXTURE. It is pretty clear, upon considering all the different kinds of chairs there are, that the resultant vector fingerprints are going to be very diverse. Consider some of the different objects that are chairs: a wooden kitchen chair, a large leather upholstered chair, a plastic lawn chair, art deco chairs, a dentist’s chair, etc. What shape do all of these chairs have in common? One might suggest that they all have an “L” shape in common, since all chairs (but not stools) have backs. But although the “L” shape may be a part of the chair in some sense, not all chairs are merely “L” shaped. But, one might continue to object, at least for something to be a chair it must have the “L” shape as a constituent, perhaps along with legs (and the legs themselves are to be specified not as the concept LEGS, since this would fall back into a definition using concepts, but, rather, as a certain kind of SHAPE).

\textsuperscript{18} Perhaps one could weight certain dimensions and this weighting of certain dimensions (i.e. the sound dimension) would be able to exclude the dog from the cat category (Smith and Heise, 1992 have an idea similar to this). But if this is so, then whether or not the dog barks becomes closer to definitive of the category DOG, in which case non-barking dogs, such as Basenjis, would not be included in the category DOG.
There are several responses to this. First, it is not at all clear that the chairs should cluster together in terms of their “L” shaped constituents. Why should being “L” shaped be that part of its shape that is most relevant to the clustering in vector space? Why wouldn’t it be the overall shape rather than a portion of its overall shape? To the extent that we can already focus on a particular portion of the overall shape, one might worry that we’re already importing some conceptual understanding in, such as that chairs are used for SITTING. Second, and more crucially, even if we were to accept that the “L” shaped portion of a chair were the key to a chair being grouped together, then it seems there would be many other things, including desks and sofas that would also be classed as chairs. Consider, for example, the similarity between a wooden chair and a wooden desk with a back on it. Or simply two pieces of wood nailed together at a 90 degree angle. In many ways, the wooden desk or the two pieces of wood are more similar in terms of SHAPE and TEXTURE to the wooden chair than an upholstered leather chair is to the wooden chair. But if so, the desk would be included in the category CHAIR. Likewise, a wooden bench is more similar to the wooden chair than the leather upholstered chair. This is just another instance of the general problem that categories formed by comparing instances in terms of the most general dimensions cannot accurately capture kinds.

The second horn of the dilemma is that if we try to build in more precise features in order that the vector clusters will accurately capture kinds, then we are coming dangerously close to presupposing concepts. Thus, we can’t assume that the dimensions of the vectors are things like WINGS and LEGS, since these are themselves concepts. Rather they must be dimensions which could plausibly characterize an innately given
vector space—dimensions such as EDGES, SHAPE, SOUND and COLOR. Again, dimensions do not themselves carve up the set of objects in the world into subsets, rather, on this view, the clusters of vectors are supposed to do that. Thus, the dimensions themselves are not concepts since they are doing no “carving up” of the world into sets of objects. However, if one were to attempt to build in dimensions like WINGS or LEGS, then these would carve up the world, since WINGS, unlike SHAPE, applies to only some objects. But in that case we would have to either explain where these concepts came from (thus entering on a regress) or accept that they’re innate. Thus, putting the two horns together, the dilemma is that one either stays true to a kosher empiricist account but can’t account for the kind concepts we have, or one presupposes concepts and either accepts a regress or accepts them as innate, abandoning one’s empiricist principles.

Given this problem, someone who wanted to defend a version of this account of abstraction might attempt to say how we could “tidy up” the clusters in vector space. If the problem is that grouping individuals according to their particular vector “fingerprints” doesn’t get us the kinds we need, then perhaps if we added more dimensions to the vectors, we could fix the interrelated problems of the vector clusters being both too wide and too narrow. I will examine a couple of suggestions about what could be added to do some work: function and language.

First one might think that if we added the function of a thing to its vector fingerprint, then we would be able to rule out some gerrymandered categories. For example, what all chairs have in common, it might be said, is not merely that they have a certain shape (since this won’t exclude some desks, boards nailed together at a 90 degree angle, and some benches), but that they have a certain function, namely sitting. But how
are we to flesh out the notion of sitting in a way that doesn’t already presuppose the
concept SITTING? Well, perhaps sitting has a particular shape profile when you
combine the things that sit (typically human beings) with the shape that is taken when
sitting. But this only captures an “image-like” feature of some instances of seeing chairs,
namely, that a person is sitting in the chair, and it is this image that I’m saying has a
particular “shape profile” which could be captured by a vector. But the image-like
feature of someone sitting in a chair is not the same as the function of sitting. Indeed, it
is difficult to see how one could flesh out the function of sitting as tied to only chairs
(and not desks) on this method, since some people do sit on desks and (of course) on
benches.  

What is needed, if this attempt is to work is a notion of function which is
something more like “what the thing is intended for,” and that’s a notion of function that
is difficult to imagine could be captured by only storing instances of chairs. What is
needed to distinguish the wooden chair from the wooden desk is the proposition “chairs
are designed for sitting on” to be attached to only the chair vectors of the vector cluster in
order to “weed out” the instances of the desk vectors, which don’t belong. But how are
we to capture the proposition “chairs are designed for sitting on” in a vector space which
merely stores instances of particulars encountered according to the general dimensions
mentioned? We can’t account for the occurrence of CHAIRS in the proposition since
that is precisely what we are trying to give an account of; but neither can we capture the

---

19 Moreover, sometimes chairs are seen without people sitting in them, in which case the
particular “shape profile” of someone sitting on the chair (which is supposed to
distinguish chairs from desks) wouldn’t be there.
concepts SITTING or DESIGNED, since these are themselves concepts no easier to account for than CHAIR.

Thus, the problem with attempting to invoke function as a dimension of a vector space in order to tidy up the (would-be) categories formed by the vector clusters is that there’s no way to represent function without invoking further concepts, which cannot be accounted for in the vector space. In addition, a more general problem with the idea of invoking function to fix what’s included in the vector cluster is that not everything has a function. What is the function of a dog? Likewise, with many other non-artifacts (to adapt a phrase from Sartre) their being precedes their function. Thus, the method of invoking function to tidy up the kind representations doesn’t generalize to many of the natural things of which we have concepts.

A defender of abstraction might also attempt to invoke language as something which narrows down the space of possible groupings to just the right grouping that corresponds to a kind. The idea here would be that some things are called “chair” while other things are called “desk,” “bench,” “sofa,” etc, and that we use the different sounds of the different names to help distinguish between chairs and desks. The problem with this suggestion, is that if words are viewed as merely another piece of information to be fed into an associative engine, then they will still be insufficient to tidy up the clusters. An account of how words could help the problem of abstraction might be able to work if one made a few assumptions. For example, if 1) one were to hear the sound of the word “chair” only when a chair is present (and not when benches or desks or present) and 2) if the sound were stored as part of the vector fingerprint and 3) if the sound were weighted in the vector such that it made those objects not occurring with that sound much less
similar than to those objects which were attached with the sound, then perhaps one’s vector cluster would include no benches or desks. Let me grant, for the sake of the argument that the second two assumptions are met; the problem is with the first assumption. It’s clearly not the case that every time someone sees a chair, they hear the word “chair,” in which case those chairs which weren’t accompanied by the word “chair” wouldn’t be stored within the vector cluster of those that were. This introduces a new cluster problem, viz. there will now be a cluster of only chair vectors, but not all chairs since some of the chair vectors will be stored in their own little cluster of objects (some chairs, some desks, some benches) which weren’t accompanied by the word “chair.”

So it couldn’t be that words help the abstraction process in this overly simplistic way. Perhaps, then, one needn’t hear the occurrence of the word every time one sees the object. But, in this case, would the vector cluster be able to be winnowed down to include all and only chairs? If one were to say that the sound becomes associated with the concept CHAIR, this already assumes we have some way to represent the kind in question, in which case, what good is the word doing? But neither can one say that the word becomes associated with certain members of the kind and then is extended, perhaps on some kind of inductive principle, to the whole kind, since this also assumes a way of representing members of a kind. For the same reason, the defender of abstraction cannot appeal to the meanings of words (as opposed to their sound) to help solve the problem of abstraction. If one appealed to the meaning of the word “chair” then one would already be presupposing a concept CHAIR. The problem is that in order to make the vector

---

20 That the cluster of objects would include some desks and benches as well as some chairs would be true if we made the same assumptions about the formations of concepts DESK and BENCH as we did for CHAIR, since not all desks or benches one would encounter would include the sound of the words “desk” or “bench.”
cluster include only chairs, one would have to come up with some property which all and only chairs had (and benches and desks didn’t). But words (as sounds) aren’t going to be very good candidates for such a property since the sound “chair” will not co-occur with all the chairs one sees (and mutatis mutandis for other words and other objects).

On the other hand, if words are supposed to play a more substantive role in the generation of concepts then there’s no hope of giving an account of communication and meaning in terms of mental representations since language would be needed to have a concept. But what the mental representationalist wants is to be able to give an account of thought in terms of mental representations, so that words turn out to express thoughts (Gauker, 1994, 2003). But if language is necessary for the formation of concepts (and thus thought) then we can no longer give an account of thinking in terms of mental representations since an account of mental representation would require a prior understanding of language. Of course, if one claimed that the account of understanding language was itself to be fleshed out in terms of mental representations, then one would enter on an infinite regress, since one would then have the same problem (of explaining how words like “chair” capture a kind) all over again. In other words, if we suppose that the child learns the concept CHAIR by being guided by an adult’s use of the word “chair,” then we are left with the problem of explaining how the adult got their concept CHAIR. Appeal to that adult’s learning the concept CHAIR in the same way quickly enters us on an infinite regress. But if the defender of abstraction, in order to avoid the infinite regress, accepts that language is in some sense constitutive of concepts, then there is no longer a purely representational theory of mind in the offing. Some such account
might be true, but it would not be an account of how mental representations could come
to stand for kinds. And this would make Granny very sad.

**Nonconceptual Thought**

One way to define nonconceptual thought is negatively, in which case it would be
something that qualifies as thinking but which cannot be characterized as making
judgments. Thus, if there is a kind of thing which qualifies as thinking but which doesn’t
subsume a particular under a kind, then it will be nonconceptual thinking. If the
foregoing is correct, then there may be some kinds of categorization behavior which
count as thinking. The problem is to give some sort of rough and ready characterization
of what counts as thinking. I won’t attempt to give a definition of thinking in terms of
necessary and sufficient conditions (not the least of which because there probably isn’t
one); rather, I will note several different kinds of behavior which are more or less
thoughtful.

Both connectionist networks and nonlinguistic animals are at the very least,
capable of learning to detect certain patterns in the environment. One characteristic of
such learning (so-called associative learning) is that it takes many, many trials of training
with the stimuli for the organism to learn the distinction at hand. This is something
which animals and artificial neural networks have in common. Just as it would take a
pigeon many, many trials to learn some distinction, so also it takes the connectionist
network many trials to learn the intended distinction—in both cases it is a slow, plodding
There are some animals, however, that can do something beyond associative learning, so-called one-shot learning. Consider the case of a chimpanzee who, when left in a room in which a bundle of bananas are hanging from the ceiling, figures out how to stack boxes in the room so that he can reach them (Köhler, 1976). This kind of thoughtful behavior is distinguished from cases of associative learning because it is creative and not-by-rote. However, it also differs from mere categorization behavior via associative learning since it doesn’t obviously involve categorizing anything. Rather, it seems to involve recognizing that some object can perform some function in the service of a goal. Another example would be something like learning how to open a latch on a door to a room which contained food. In both of these examples, there is clearly something that the chimp wants to do, something that it’s trying to accomplish. And this is shown by the novelty, complexity and persistence of the behaviors in which the creature engages.

For a case a bit closer to home, think of some of the behaviors in which humans engage which are obviously thoughtful, but which don’t obviously involve the exercising of concepts. The best examples of this kind of thinking are different cases of “know how.” For example, consider the ability of a rock climber to find a route up a wall. This is a kind of engaged and concentrated behavior but it doesn’t seem to require that one be making any judgments. Rather, it involves something more like the ability to recognize affordances in one’s environment (in this case, where one should place ones foot, how one could get to the next hand hold, etc.). It is often noted that animals see their environment in terms of affordances for action. For example, the bird might see certain

---

21 Except that in the case of the network we can speed up the training process by super fast computers, so that perhaps it doesn’t look quite as slow.
places as a good place to build a nest. But the bird need not be making any judgments to the effect of “this is a good place to build a nest.” Likewise, it seems, even though the climber’s behavior seems to be thoughtful, it need not involve making any judgments like “that hand hold is a good one.”

It seems clear, then, that there is thinking which doesn’t involve exercising concepts in judgments. The more difficult task, however, is to give a positive characterization of nonconceptual thought. In the next chapter I will present a possible way of thinking about nonconceptual thought. Fundamental to this account will be the idea of a similarity space: an n-dimensional qualitative space by which a creature could categorize X as more like Y than Z without thereby having the ability to represent X as particular kind of thing, e.g. a DOG.
CHAPTER 2: EMPIRICIST CONCEPTS

How are we to define “concept?” Many people could be made to agree that it’s a mental representation which captures some understanding of a kind of thing. That is, it stands for more than one object. For example, having the concept DOG means that I can represent not just this dog but dogs in general. But what does it mean to represent “dogs in general?” In chapter one I argued that being able to represent “dogs in general” means being able to think of a particular (e.g. this object) as belonging to a kind (DOG). Thus having the concept DOG entails at the very least being able to think a predicative thought like X IS A DOG. Such predicative thoughts I call judgments. Thus, my claim is that if you have a concept then you can make judgments using that concept.

Some psychologists think that certain categorizing behaviors of prelinguistic infants are to be understood as revealing that children have concepts. For example, infants who look longer at an unexpected (from the adult point of view) outcome are inferred to have certain expectations concerning objects and these expectations concerning objects are taken to be constitutive of their “concepts.” Likewise, if children look longer at a cat after seeing a bunch of dogs, this preference for the cat is taken to show that they see the cat as different from the dogs and thus can be said to have a concept DOG (see, e.g. Quinn 2002; Quinn and Eimas 2000). (And if the child didn’t distinguish between dogs and cats but did distinguish between dogs and chairs or cats and tractors, then perhaps they have the concept pair ANIMAL and INANIMATE.) But are these infants making judgments? Are they thinking thoughts like X IS A DOG? I would guess that many psychologists who maintain the children have concepts would
nonetheless deny that children are thinking such thoughts. Nevertheless, they would continue to maintain that children have concepts in virtue of the fact that children have expectations concerning the behaviors of physical objects (object concept) and that they are able to categorize certain kinds of objects (kind concepts).

However, if it is correct to say that to have a concept is to be able to make judgments using that concept, then it cannot be correct to say both that children do not make judgments and that they have concepts. At this point, someone who wanted to continue to maintain that prelingustic infants have concepts could either argue that categorizing behavior is in fact indicative of judgments being made (perhaps in a language of thought) or that “concept” should not be defined using “ability to make judgments” as a necessary condition. In chapter one I argued that categorization and judgment differ in significant ways. In this chapter I will argue that children’s categorizing behaviors need not be understood in terms of making judgments (i.e. predicating a kind of a particular) since they can be explained by positing an ability to make similarity comparisons which themselves do not invoke representations of kinds. Moreover, the abilities the child displays do not fit the characterization of conceptual thought presented in chapter one.

One might object that the fact that one can explain categorization without positing judgments doesn’t imply that children don’t make judgments. I guess that could be the case, but if it is already antecedently unlikely that prelingustic infants are thinking thoughts like THAT IS AN ANIMAL or THAT IS A DOG, then an alternative account which undercuts the main reason for maintaining that they do should seriously weaken the case that they do. Perhaps a response to this would be that infants represent kinds
“implicitly” without making the explicit judgment that, e.g. THAT IS A DOG. Although this is a possibility, it’s not clear why we would need to posit a further cognitive operation if we were able to account for categorization in terms of other cognitive operations.

On the other hand, maintaining that judgment is not constitutive of concept possession seems to abandon the very heart of what it means to possess a concept inasmuch as concept possession is about the ability to represent kinds. Perhaps some would say that categorization (rather than judgment) is the most central ability relevant to concept possession. But although someone possessing a concept would be able to categorize, a common way of understanding how a concept enables one to do so is by judging whether a given particular falls under the concept. Thus, categorization would be parasitic on the more basic ability to form judgments. However, I have claimed the converse is not true: the ability to categorize (for at least some kinds of categorization) does not entail having kind concepts.

There is a further problem with the view that prelinguistic infants have concepts. That is that there is no good account of how they could come to represent kinds. This is a problem with the mechanism by which kind representations are constructed. If one is to attempt to give an account of how kind concepts are constructed (as opposed to admitting that they’re innate) then one will encounter the problem of abstraction, which is the problem of explaining why one concept is created as the result of experience rather than another (as a result of that same experience).

I have two main goals for this chapter. The first is to flesh out the problem of abstraction using a couple recent empiricist accounts (Mandler, 2004; Barsalou, 1999) as
targets. The second is to suggest a way of understanding infant categorizing which doesn’t rely on attributing concepts (kind representations). Understanding prelinguistic thought will still involve attributing some sort of representation, albeit a representation that is nonconceptual in that it doesn’t allow for the representation of particulars as belonging to kinds. The gist of my answer will be that they do so by comparing objects within a “similarity space,” which itself cannot be understood as constitutive of representations of kinds. If I can make this case, then we can understand how categorization could take place without having to posit kind representations. And that is a good thing since, if my critique of accounts of abstraction holds, we should be wary of attributing kind representations to prelinguistic creatures.

**Infant categorization in Mandler’s work**

Jean Mandler (2004) presents an interesting and comprehensive theory of conceptual development in childhood, the main tenets of which are the following: 1) an infant’s first concepts are global in nature, 2) conceptual development follows a path from global to specific, 3) infants have two different kinds of information processing systems right from the start—one the “contents” of which are perceptual, cannot be brought to conscious awareness, and are processed implicitly, and another the contents of which are conceptual, can be brought to conscious awareness, and are processed explicitly (i.e., require attention to be processed). Thus, Mandler believes that there is such a thing as prelinguistic conceptual representation. The main reason Mandler (2004) posits prelinguistic conceptual thought is to account for the categorizing behaviors of prelinguistic infants. In particular, she thinks that some categorization behaviors of
infants are based on the representation of kinds and cannot be derived from mere perceptual similarity. One of the findings from Mandler’s work is that infants group objects together not in terms of perceptual similarities such as color and shape, but in terms of their “defining features.” Defining features, Mandler argues, are deeper than surface level perceptual features and are derived from the way objects move through space as well as their spatial relationships to each other. For example, an infant might treat a frying pan and a teacup as members of the same kind, container. According to Mandler, the principles according to which infants group together perceptually dissimilar objects supports the idea that they have inchoate concepts and are not operating merely on perceptual similarity.

One line of support comes from the object-examination test. The basic idea is a variation of the habituation-dishabituation paradigm: an infant is “habituated” to a series of objects of a certain kind, for example, different dog figurines. Habituation occurs when the infant loses interest in the object (as measured by decreased handling time). Then the infant is presented with an exemplar of a novel kind, e.g., a fish. The dependent variable is whether or not the infant “dishabituates” (handling time increases again) to the new stimulus. The rationale behind this paradigm is that infants will handle longer an object they consider novel. So, for example, if, after habituating to a series of dog figurines, infants dishabituate to the fish (i.e., handle it longer) then this would show us that they consider the object to be different from the others. In short, it will show us how they categorize. Now although it might have turned out that infants categorize only in
terms of perceptual similarity such as color or shape, this seems not to be the case. Instead, they typically fail to distinguish between fish and dogs but succeed at distinguishing animals from vehicles (2004, pp. 153-155, 168). Moreover, they will distinguish between birds and airplanes even though the models used in the experiment are very similar in overall shape (ibid., p. 155). Mandler argues that this data supports her contention that infants have concepts such as ANIMAL (characterized as “self motion, biological motion, and contingent interaction with others at a distance”) and it is this that accounts for the categorization distinction between animals and vehicles, which, it seems, are classes which differ widely in the perceptual similarity of their members.

According to Mandler (2004, p. 10), the most conclusive evidence for prelinguistic concepts comes from the generalized imitation paradigm. Here, the experimenter will act out some scene with life-like figurines, e.g., giving a figurine of a person a drink with a teacup, and then the infant is given a number of exemplars of different kinds. The question is to what extent the infant will act out this scene with exemplars from different categories. For example, will the infant give a drink to a car? Will it give a drink to the figurine of the person using a frying pan instead of a coffee mug? The rationale behind this paradigm is that the kinds of objects infants will

---

22 It should be noted here that children do categorize according to color or shape in other looking time experiments where the child is shown pictures of animals rather than given figurines to handle (Mandler, 2004, pp. 161-162). Mandler suggests that the reason that children respond to perceptual differences in the pictures task but not the object handling task is that children are much more interested in (and so attend more to) the objects than the pictures and that this increased attention activates conceptual processing in the object handling task but not in the picture looking tasks (2004, pp. 151, 162, 163). Mandler thinks that we needn’t posit concepts to explain the categorization by color, shape, etc. (2004, p. 152).
generalize to tell us about how they understand the event, what things they consider to be of the same kind (2004, p. 11).

Here is one interesting result: when the experimenter gives a (figurine of a) person a drink with a teacup, infants are just as likely to imitate this action using a frying pan as using a coffee mug. What this shows, according to Mandler, is that infants understand the cup in the modeled event not so much as a CUP but as something broader like CONTAINER. This is a case of infants failing to categorize at the basic level in lieu of a more superordinate categorization, such as container. Another example of this kind is that they will only imitate the action of giving a drink with animals, not vehicles; and they are just as likely to use any other animal as the one that was used to model the action. Mandler’s contention is that the understanding that underlies categorizing something as a container or an animal must be conceptual in nature. The reason she thinks this is that infants seem to be categorizing objects according to what the objects do and how they interact, and the basis of this kind of categorization can only be concepts (which abstract from perceptual similarity).

The explanans and the problem of abstraction

Mandler’s main theoretical entity is what she calls an “image-schema.” Image-schemas are representations that summarize “spatial relations and movement in space” and which “express primitive or fundamental meanings” (2004, pp. 78-79). Although Mandler doesn’t use the terminology, she is viewing image-schemas as having conceptual
Image-schemas are “abstracted” from perceptual information via what Mandler calls “perceptual meaning analysis,” which occurs as the result of noticing or paying attention to different aspects of motion events (where the tendency to pay attention to motion is, perhaps, innately specified).

The simplest image-schema is that of PATH, which is just the trajectory of some object through space without any details of the trajectory or of the object. If, however, attention were paid to particular aspects of the PATH, then other image-schemas would be abstracted. For example, if the onset of motion is attended to, then either the SELF-MOTION or CAUSED-MOTION image-schema will be formed depending on whether the object began to move on its own or came into contact with another object. If the manner of the PATH is attended to then the infant will form the image-schema ANIMATE MOTION, which is something like “rhythmic, up and down, irregular” (2004, p. 96). In addition, there are several LINK schemas that represent contingency relations between objects. A one-way link represents regularities: if A happens then B happens; a two-way link summarizes various kinds of give and take interactions.

Combining these image-schemas, we get a picture of how the infant might conceptualize animals as “things that start themselves, move in a rhythmic although not always predictable way, and interact with other objects contingently both directly and from a distance” (2004, p. 99). Mandler believes that in this way infants acquire global kind concepts such as ANIMAL and that it is this kind of conceptual representation that accounts for infants’ categorizing behavior in paradigms recounted above.

---

23 See (2004) p. 70: “meanings that are used to create concepts such as apple”; p. 81: “permanent representations of meaning”; p. 14: “perceptual meaning analysis extracts and summarizes a subset of incoming perceptual information from which it creates a store of meaning or simple concepts”
For Mandler, the virtue of an image-schema is that it stores merely the spatial information and abstracts away from all the figural details. However, her account of how this happens pushes her headlong into the problem of abstraction.

**Mandler’s account of abstraction**

Mandler wants to be a concept empiricist, which means that she wants to show how concepts can be created out of nonconceptual representational resources. However, she is willing to admit (as would be any empiricist) that we come built with certain predispositions to respond differentially to different kinds of stimuli. The concept empiricist can freely admit such innate biases so long as they aren’t conceptual in nature.

For example, Mandler thinks that the so-called “object concept” in the developmental literature (see chapter 3) is not so much a concept as it is an “ontological commitment” (2004, p. 63). Unlike the concept DOG or ANIMAL, which can be explicitly called to mind and reasoned about, an ontological commitment is implicit and procedural (ibid., p. 64). Perhaps certain commitments such as that solid objects cannot pass through each other are innate but they shouldn’t be viewed as the same kind of things as our concepts of dogs and chairs. One could admit innate ontological commitments while maintaining that concepts of dogs and chairs are learned. So Mandler’s task is to explain how concepts such as ANIMAL are created without positing innate concepts (but perhaps innate “commitments”).

It is important to note here that the basis on which Mandler wishes to draw a distinction between “ontological commitments” and the “concepts of more specific objects” is not as obvious as she might wish. She thinks that the object concept is more
general, not concepts of specific things (ibid., p. 63). But in what sense is the concept ANIMAL of specific things while OBJECT or SPELKE-OBJECT isn’t? The infant’s object concept is about specific things too, just a broader range of things, i.e. “unified, bounded, persisting bodies” (Spelke, 1990, p. 31). So Mandler can’t say that there is some difference in kind between concepts of Spelke objects or concepts of animals in terms of their being concepts of specific things. Instead, she must make the distinction, as I assume she would, in terms of ability to consciously think and reason using them. The infant’s object concept is not something that fits that bill since (presumably) it is not a representation they can call to mind.

As noted in chapter one, there seem to be two different ways the mind could create concepts from percepts: either by comparing similarities and differences of percepts (comparative view) or by creating the concept from just one percept (noncomparative view). Mandler seems to accept both views as an explanation of how percepts can be turned into concepts (2004, pp. 68-70, 73). In explaining how infants acquire their first concepts, however, she accepts the noncomparative view:

Notice that although conceptualizations of animate and inanimate things result from perceptual meaning analysis, the mechanism itself does not require comparison of one kind of motion with another. The baby can analyze self-motion, for instance, simply by observing that an object began to move in the absence of any other object touching it (2004, p. 73).

At least part of the reason that Mandler would seem to hold the noncomparative view of abstraction is that she doesn’t think that percepts are consciously accessible since, on her view, the only things that can be stored in such a way so as to be consciously accessible at a later time are concepts (2004, pp. 21, 222). Thus, not yet having any concepts, the infant wouldn’t have any consciously accessible representations which it could call to
mind to compare with current perception. However, I don’t think that there’s any good reason to say that the only things that can be consciously accessible at a later time are concepts. Why couldn’t we consciously recall stored images? Mandler’s own argument certainly doesn’t establish that it is not possible, as I will now show.

Mandler’s evidence to show that concepts are required for conscious recall is an experiment by Carmichael, Hogen, and Walter (1932). The experiment consisted of showing people nonsense drawings with different labels and then having them reproduce the drawings at a later time. For example, a drawing of a line with a circle on each end was labeled as either a dumbbell or eyeglasses. When the subjects were asked to reproduce the figure at a later time, the reproductions took on further characteristics of either dumbbells or eyeglasses that weren’t in the original figure. So those subjects for whom the figure had been labeled “dumbbell” drew dumbbell-like features that weren’t in the original figure and subjects for whom the figure was labeled “eyeglasses” drew eyeglass-like features.

It is not surprising that conceptualization influences imagistic memory, but this fact hardly establishes that conceptualization is necessary to recall past events, it only shows that it can be. This experiment establishes neither that an adult’s recall of images must be conceptually mediated, nor that an infant’s recall must be mediated by preverbal concepts. There’s no reason to conclude, on the basis of this data alone, that conceptual encoding is necessary to recall, say, an image of the bright yellow tree I saw while running through the forest yesterday. Rather, it seems like such experiments would only establish the following conditional: “If we have conceptualized two circles connected by
a line as eyeglasses, then something that looks like eyeglasses is what we will recall.”  

But this is no reason to think that recall of images require conceptualization.

Be that as it may, Mandler is committed to the noncomparative view of abstraction in the case of an infant’s first concepts. So how does the noncomparative view work? Mandler’s mechanism for abstraction is what she calls “perceptual meaning analysis.” Perceptual meaning analysis creates from perceptual displays “image-schemas,” which are “analogue representations that summarize spatial relations and movements in space” (2004, p. 79) and are consciously accessible. Image-schemas are concepts; they are consciously accessible and although not atomic “may serve the same function as words in infants’ thought” (ibid.). As such, “they could, if one had some reason to want to, be combined into propositions” (ibid.). Passages such as these strongly suggest that Mandler views image-schemas as able to fill the role of concepts as constituents of thought and judgment. Mandler is at pains to make clear that image schemas are not images and thinks that much misunderstanding has resulted from seeing them as such (ibid., pp. 79-80). They are representations which “eliminate figural details, including many details of movement” (p. 81). Citing Barsalou (1999), Mandler notes that it is possible to represent shape without orientation and path without a direction. But the kind of representation at issue here is perceptual, not conceptual. The fact that there are high-level neurons that represent, e.g. color without shape, doesn’t help Mandler in establishing her point, which is that there exist conceptual representations (i.e. a

24 This conditional is Mandler’s own; apparently she doesn’t realize that its truth is independent of the truth of the claim that “concepts are required to mold perceptual data into imaginal form” (2004, p. 21).

25 But contrast p. 119: “I have tried to show here that a propositional preverbal system is not necessary….Indeed it is quite possible that propositional representation simply does not exist in the human mind until language is learned.”
Barsalou himself has a theory about exactly how these high level neurons can play a role in representing kinds, but for him it is not the high-level neurons themselves that account for the ability to represent kinds, it is his “simulators.” But it is not enough for Mandler to claim that the visual system processes information in a particular way, since she’s attempting to make a claim about how concepts are formed, not how the visual system processes information about objects. Of course, facts about the visual system could be relevant to an account of concepts, but that would have to be developed. Merely citing facts about the visual system as evidence for the existence of certain “abstract” concepts works about as well as claiming that our concept EDGE comes from the edge detection mechanism in our low-level visual system.

How, then, are image-schemas created? It is here that Mandler’s account becomes mysterious as to just what perceptual meaning analysis is supposed to do. Here’s what’s supposed to happen: being drawn to attend to moving things, the infant’s perceptual meaning analysis (PMA) mechanism then turns the perceptual information into an image-schema. For example, the infant sees a moving object and attending to it abstracts the image-schema SELF MOTION. But now we ask the question, why does the infant abstract SELF MOTION rather than MOTION AWAY FROM ME or MOTION TOWARDS ME or some other thing? How does PMA arrive at SELF MOTION? Unless she has some answer to why X is abstracted rather than Y or Z, the PMA mechanism is a *deus ex machina.*

---

26 Barsalou’s theory, however, has its own problems, which I present below.
Mandler seems to realize that what PMA abstracts must somehow be constrained since there is more than one thing that could be abstracted: “Nevertheless, there must be a core of possible meanings that human creatures can extract by means of this mechanism. Why is this meaning extracted rather than that meaning” (2004, p. 67; c.f. pp. 83-84)? However, she underestimates the force of her own question. She seems to think that the “perceptual system’s ability to discriminate motion parameters” enables us to get some of the image-schemas for free (ibid., p. 71). Mandler says that our ability to perceive biological vs. nonbiological motion, contingent motion and self vs. other-instigated motion “is a simple perceptual discrimination” (ibid., p. 72). But although we can no doubt see, for example, an object which starts up on its own, we don’t just automatically thereby conceptualize it as a self-moving object. So too for infants’ differentiation between animate and inanimate objects. Infants may well be able to respond differentially to animate vs. inanimate motion (as Mandler suggests, p. 71) but this is procedural knowledge (in Mandler’s terms) and it is another step to say how perceptual sensitivities can be turned into the concepts/image-schemas ANIMATE MOTION and INANIMATE MOTION. Analogously, although our perceptual system has the ability to distinguish between black and white, the concepts BLACK and WHITE do not just come for free as a result of our perceptual system’s sensitivity to such parameters. But perhaps what she means is that that we have to ability to attend to particular aspects of movement and that what we attend to is what PMA abstracts (“attention sets the stage for analysis” (p. 72)). But even if we were innately attuned to pay attention to, e.g., whether an object moves on its own or is contacted by another object, there are any number of different ways to

27 Perhaps Mandler could admit some image schemas are innate, something which in later work she clearly does. I examine this option in more detail below.
conceptualize what we are attending to. What we attend to does not determine the concepts we abstract.

One of the most interesting aspects of Mandler’s view is the idea that the kind of environmental information that infants are most sensitive to is spatial and movement information (2004, pp. 72, 78) rather than featural detail of objects. There is a bias to pay attention to moving objects (ibid., p. 301). Mandler speculates that although in the infant’s first month they are not yet receiving “high-quality foveal information about what the objects moving around them look like,” they can still get good information about how objects move (ibid., pp. 70-71) and that perhaps these developmental constraints are the cause of the infant’s focus on spatial and movement information. However, although the immature visual system could be part of the reason that infants attend to movement information, it cannot be this alone which causes them to focus specifically on, say, the path of the motion or on self-motion. That is too broad of a constraint to get you to such a specific thing to focus one’s attention on. Mandler could admit that there is an innate bias to pay attention to specific kinds of movement such as self-motion.28 What she cannot say (without thereby giving up her empiricist aspirations) is that many or most of the concepts/image-schemas which the PMA mechanism creates from certain perceptual displays are innately determined. For in that case, what the mechanism is to abstract is already innately determined rather than created from the perceptual information. It is also fine for Mandler to claim that the mechanism of perceptual meaning analysis is itself innate (analogous to the way that mechanisms for associative learning are innate in animals) (ibid., pp. 66, 300), but then she must explain

28 Although even here there’s the worry that being able to attend to such specific things like self-motion already presupposes such concepts.
how the mechanism does what it does without being innately determined to abstract just those concepts she needs, i.e. SELF-MOTION, ANIMATE MOTION, PATH, etc.

Unfortunately, she doesn’t do this. As I have attempted to show above, Mandler’s account does not answer the problem of abstraction. Our (possibly innate) ability to perceive motion parameters or to attend to those motion parameters is insufficient to determine the particular concepts Mandler thinks we abstract. At one point she even admits that which image-schemas are abstracted from a perceptual display must be innately determined:

One such bias, I have speculated, is a tendency to attend to the beginnings and ends of paths. Such biases may be characteristics of the perceptual system itself and not unique to perceptual meaning analysis. However once attention is drawn to the start of motion or to its cessation, the occasion for perceptual meaning analysis exists and…leads to concepts such as agent and goal. It may also be, as in the case of attending to object paths, that these biases are not just side effects of a particular kind of perceptual system. To conceive of a self-starting object as an agent or the end of a path as a goal cannot be characteristics of the perceptual system itself. Hence, they may represent innate biases of perceptual meaning analysis to interpret paths in a certain way (2004, p. 301, my emphasis, cf. p. 84).

But for Mandler to admit this is to give up the empiricist project of deriving concepts from percepts, is it not?29 The problem is that the result Mandler admits here is not particular to the image-schemas AGENT and GOAL. Even the simplest image-schema, PATH, is problematic since it is no more a “characteristic of the perceptual system” than AGENT or GOAL. PATH is supposed to represent the trajectory of an object without specifying its “speed, direction, jerkiness or smoothness, or any other figural aspect” (p. 84). It’s one thing to say that we can attend to the path of the object and another thing to say that we can conceptualize the path as such. There’s a gap between the “attending to”

29 Perhaps the empiricist can admit a small stock of innate concepts and then go on to show how the rest are acquired from experience. I examine this idea below.
and the “conceiving as” into which we can insert the question: why extract this meaning rather than that meaning? The attending to does not determine the concept to be abstracted.

More recently, Mandler has even more explicitly admitted (forthcoming; cf. 2004, p. 301) that PMA innately specifies what to abstract for a number of the spatial primitives including PATH, LINK and ANIMATE MOTION. But in this case, PMA is not creating concepts out of percepts, rather, the concepts are already innately specified. Does PMA create image-schemas which aren’t innately specified? Mandler seems to think so since she thinks that any distinctions a natural language can make must be represented by image-schemas (2004, pp. 118, 251) and these distinctions come to be represented by a process of differentiation from those built up from primitives (2004, p. 202; forthcoming). Thus, although she doesn’t face the problem of abstraction for the innate primitives (or those that can be built from them), she’ll still have to explain how the image-schemas which aren’t innate are created. Perhaps PMA will abstract whatever it is we attend to, which is the only other account I can glean from Mandler (2004, pp. 72, 257). But innate attentional capacities are insufficient to determine the concepts abstracted because there’s always a gap between what we attend to and how we are to conceptualize what we are attending to. We’re back to the same problem: PMA does not have the resources to abstract the particular concepts that need to be abstracted without some kind of already specified recipe of what to abstract. Claiming that PMA always follows an already specified recipe is tantamount to the claim that all concepts are innate. Although some philosophers have bitten this bullet, I’m not sure it’s one that Mandler would like to bite.
Mandler (forthcoming) discusses two other ways of forming concepts: language and analogy. However, as she admits, language merely directs attention to different phenomena, it doesn’t account for the underlying representations. In the case of analogy, she thinks that our existing spatial concepts can point to various sensorimotor experiences, e.g. time is represented by spatial concepts. But it’s unclear how exactly the spatial image-schemas could combine with sensorimotor experience (something which, on Mandler’s own account, is purely procedural and not accessible to consciousness) to make a new conceptual representation.

**Interlude: Is conceptual representation processed implicitly?**

Fundamental to Mandler’s view is a distinction between two different kinds of representational format: perceptual and conceptual. These two representational formats differ in representational format and methods of processing (2004, p. 45). The intuitive distinction she wants to draw is between “knowing how” and “knowing that” (Ryle 1949). “Know how” is characteristically exhibited in certain skills while “knowledge that” is exhibited by certain statements. Contrast one’s ability to ride a bicycle with one’s knowledge of who invented the first bicycle. If I wanted to find out whether your claim that you know how to ride a bicycle is true, then I would ask you to ride the bicycle. If you get on the bicycle and then can’t keep your balance and keep wobbling and tipping over, then you will have shown your statement “I know how to ride a bicycle” to be false. In this case, I discover that your statement is false because of the behavior you exhibit, but this behavior itself is not true or false. On the other hand, if I wanted to find out if you knew who invented the first bicycle, then I’d simply ask you. If you replied that
Lance Armstrong invented the first bicycle then your statement “I know who invented the first bicycle” is false. In this latter case, however, I discover that your statement is false because of the falsity of another statement (viz., that Lance Armstrong invented the first bicycle).

Whereas I can exhibit knowledge of how to ride a bicycle, I cannot describe to you how I accomplish this feat since it is not something of which I am aware in the exercise of my “know how.” I may be able to make up stories about how I ride the bike (for example, I may say that I feel myself leaning to the right to compensate and lean back to the left) but this is different from being able to explain the processes which actually enable me to accomplish this feat, e.g. the sensory-motor machinery my body uses, such as efference copies, to balance itself on the bicycle. Someone might object that, in fact, I could put into words how I ride a bicycle, if only I would take a physiology class and learn about the sensory-motor representations my body uses to balance itself on the bicycle. In that case, I could put into words what I am doing. Although one could do this, I claim that it would no longer count as “knowledge how” since a person could have this knowledge and yet not have the ability to ride a bicycle. If you replied to my request for you to ride the bicycle with a long, drawn-out story about the physiological mechanisms which accomplish this act, I would say “thank you for the physiology lesson, now please get on this bike and ride it.” That’s the only way I could be assured that you knew how to ride a bicycle. Of course, you might both be able to ride a bicycle and put into words the physiological story about how this happens. In this case, however, the physiological story is not something on which you draw in the exercise of your “know how.” It’s totally epiphenomenal to your knowledge of how to ride a bicycle. This
contrasts with “knowledge that” since being consciously aware of statements is (typically) part of the exhibition of one’s “knowledge that.”

If I ask you who invented the first bicycle then in your exhibition of your “knowledge that” you will be consciously aware of the very thing of which your knowledge consists, namely, the statement “Baron von Drais invented the first bicycle in 1817.”

Mandler calls “know how” procedural knowledge and calls “knowledge that” declarative knowledge. According to her, what distinguishes one from the other is whether or not the representations can be brought to conscious awareness: procedural knowledge cannot while declarative knowledge can be brought to conscious awareness. For example, the mechanisms which underlie our ability to tell the difference between male and female faces are not something we are aware of (2004, p. 48), nor is an awareness of how these mechanisms work (if one happened to have this) a part of this ability. Mandler claims that perceptual recognition (such as facial recognition) is procedural knowledge how whereas conception is declarative knowledge that (ibid, p. 46).

Closely related to the differences in representational format are differences in methods of processing. Procedural knowledge, Mandler claims, can only be processed implicitly, which is to say, one is not consciously aware of the representational mechanisms which constitute procedural knowledge nor would such an awareness (if one had it) be a part of the exhibition of procedural knowledge. Declarative knowledge, on the other hand, is processed explicitly, which is to say that one is aware of it and can describe it with language or (with some training) draw it (2004, pp. 54-55).

30 The parenthetical “typically” expresses my reservations about cases of hypnotism and other fanciful examples one might come up with.
Thus both the representational format differences and the methods of processing differences turn on whether the representations can be called to conscious awareness. So in what ways do format and processing come apart? The reason Mandler seems to want to posit a distinction here is that she thinks declarative knowledge can be processed implicitly. The evidence that Mandler gives for there being a distinction between format and processing is an experiment which consisted of subjects being given a list of words and either one of two tasks to perform on those words. One task was just to study the list of words while the other was to cross out the vowels in words on the list. There were two different testing variables: a stem completion task, in which the subject had to complete the rest of the word, given a three letter stem, with the first word that came to mind, and a recall and recognition task, in which the subject had to say whether they had seen the words before. The stem completion task was supposed to measure priming effects since the first words that come to mind depends on the most recently seen words. The interesting result is that the two groups (studying words vs. crossing out vowels) performed the same on the stem completion task but the studying-words-group significantly outperformed the crossing-out-vowels-group on the recognition and recall tasks. Mandler’s explanation of this fact is that the two groups were processing the information differently: the vowel-group was processing the words implicitly while the word-group was processing the words explicitly. Mandler seems to draw the conclusion from this that declarative knowledge can be processed implicitly (2004, p. 53). But why should we think of this as a case of declarative/conceptual representations being processed implicitly rather than just another case of procedural/perceptual representations being processed implicitly?
The priming effect is supposed to show that representations of words influence behavior even though one isn’t consciously aware of the words. Since the vowel-group shows these priming effects and yet lacks the ability to (consciously) recognize or recall words that were on the list, this is evidence to Mandler that conceptual representations can be processed implicitly. The assumption Mandler seems to be making is that the representation of words, whether accessible or not, must be conceptual in format. Let’s grant that those in the word-studying-group are representing the words conceptually since it is plausible that they are attending to what the words mean. It’s not clear, however, that those in the vowel-counting-group are representing (in whatever sense accounts for the priming effects) the words conceptually since if they (presumably) aren’t paying attention to the words as such, a fortiori they aren’t paying attention to their meanings.

For Mandler’s point to hold it must be the case that words are represented conceptually in the case of the vowel-counting-group, for in that case we would have implicit processing (i.e. not available to conscious awareness) of conceptual/declarative representation. But it’s not clear that they are in the case at hand.

Mandler says of the priming experiment that “language, even though conceptual in nature, can sometimes be processed implicitly; it is registered but not attended to or elaborated. That is, it is not processed in terms of its semantic content…” (2004, p. 52, my emphasis). What is the “it” which is “registered” and “processed” in the quote above? In this passage, it seems it has to be representations of words, since only representations are “processed”. So the question is: why should we think that the representations of words are conceptual in the case of the vowel-counting-group? Perhaps the reasoning is that the letters are grouped together in units and the only way to
see these units as units is to see them as words, and words can only be seen as words in terms of their semantic properties. But it doesn’t seem that the representations of words that enable perceptual priming in the stem completion task need be viewed as conceptual representations. What rules out their being perceptual representations? Perhaps we can group words together by their shape alone, just as a result of seeing them so many times. If so, then the representations that the vowel-counting-group which account for the priming effects needn’t be seen as conceptual or declarative representations. But then there would also be no reason for thinking that the priming effects show that conceptual/declarative representations can be processed implicitly.

The reason this point is of some significance is that Mandler claims that infants are representing objects conceptually even if these representations are not processed explicitly. The infant is essentially supposed to be making some sort of judgment like “x is a self-moving, contingent interactor” albeit not explicitly representing these thoughts. The fact that such information can be implicitly processed would seem to be an important idea for Mandler inasmuch as it seems implausible to maintain that children are explicitly processing the differences which characterize their concepts. Thus, if we question whether there’s any evidence for thinking that concepts can be processed implicitly then this would also raise questions about the plausibility of attributing conceptual thought to prelinguistic infants. Mandler’s vowel counting example doesn’t seem to establish the point she wants to make, viz. that concepts can be processed implicitly, and it is not obvious that one could make judgments such as “x is an animal” implicitly.

---

31 Of course prelinguistic infants do not represent these concepts in words, but it also seems unlikely that 7 month old infants are explicitly representing “self-moving, contingent interactor” in images either. This is why it would seem important for Mandler’s view that concept representation can be processed implicitly.
Barsalou’s account of abstraction

Barsalou (1999, 2005) offers a theory of conceptual representation based on sensorimotor reenactment. There are two stages: storage of perceptual information and simulation, which Barsalou calls a reenactment of earlier perceptual representations. A perceptual symbol is a record of a small portion of the total neural activation that arises during perception, where this subset is determined by selective attention to certain aspects of an object or event. Selective attention isolates information in perception (i.e., a portion of the active neurons underlying a perceptual state) and stores it in long-term memory. The information that is stored is thus highly schematized according to which aspects of the perceptual state are attended to. For example, if one were to attend to shape, the perceptual symbol stored would filter out color, texture, position and surrounding objects. If one were to attend to the movement of the object, then shape, orientation, color, etc. would be filtered out. The idea that such information can be isolated is supported by the neuroanatomy of vision, according to which there are distinct channels for processing different dimensions such as shape, color, orientation, movement, etc (Barsalou, 1999, p. 584). In addition, as Barsalou notes, it is well known that there exist “high-level neurons” which can encode qualitative information, such as the presence of stripes, while leaving many parameters indeterminate. For example, a high-level neuron might detect stripes on a tiger without thereby storing a determinate number of stripes, or detect the presence of a line without coding its length or orientation.

Perceptual symbols come grouped together in long-term memory in what Barsalou calls “simulators.” A simulator, which is equivalent to a concept for Barsalou,
consists of both a frame and a (potentially limitless) number of particular simulations (1999, pp. 586, 587). The frame is what collects various perceptual symbols (depending on which aspects are selectively attended to) as belonging to either the same object category or same event type. In the case of an object, the perceptual symbols would be integrated according to some sort of scheme for representing spatial relations; in the case of an event sequence, the perceptual symbols would be integrated according to some scheme for representing temporal relations. For example, consider attending to the different aspects of a car: its overall shape, the doors, the wheels, etc. Selectively attending to these different aspects creates perceptual symbols which are integrated spatially in an object-centered frame (a simulator). Contained in a frame would be many other perceptual symbols stored from other instances of cars and from this frame are constructed particular simulations. Like perceptual symbols, simulations are themselves schematic and need not involve all the kinds of information contained within the frame. So, for example, one might simulate a car without simulating its color or one might simulate the event of giving someone a drink without simulating all the features of the thing used to contain the liquid. Categorization, on Barsalou’s theory, is a matter of being able to construct a simulation for the object in question: if a “satisfactory simulation” can be constructed for the current percept then the current percept is stored in the frame/simulator from which the “satisfactory simulation” was constructed.

Barsalou gives a more streamlined account in his 2005. When perceiving an entity, various “feature maps” (neurons which act as detectors for various features of the visual array) are activated by the sensory information. For example, “during visual

32 Simulations need not be conscious either, although they can be as in the case of imagery.
processing of a chair...some neurons fire for edges, vertices, and planar surfaces, whereas others fire for color, orientation, and direction of movement” (2005, p. 398).

Next, “conjunctive neurons” capture this pattern of features and store it in an association area (this is a perceptual symbol). One thing to note at this point is that the features for which there are detectors are going to be very simple kinds of features like “edge” not complex features like “bird,” a point which will be important to bear in mind when thinking about how simulators are created. The ability to represent objects in their absence is due to the fact that the conjunctive neurons can reactivate the same pattern that was captured during perception:

While remembering a perceived object, for example, conjunctive neurons reenact the sensorimotor states that were active while encoding it. Similarly, when representing a concept, conjunctive neurons reenact the sensorimotor states characteristic of its instances (2005, p. 399, my emphasis).

“Simulator” is Barsalou’s term for concept (1999, p. 587); it is the pattern captured by the conjunctive neurons which is characteristic of multiple instances of a category:

As multiple instances of the same concept are encountered, they tend to activate similar neural states in the feature maps...As a result, similar populations of conjunctive neurons tend to capture these states...Over time, conjunctive neurons integrate sensorimotor features across diverse category instances and across diverse settings, establishing a multimodal representation of a category (2005, p. 400).

Simulators are Barsalou’s key theoretical entity, all the roles that concepts have traditionally been thought to fill are to be accounted for in terms of simulators. Barsalou is attempting to account for two properties of concepts as mental particulars: their ability to figure into subject-predicate judgments (what Barsalou calls “interpretations”) and their ability to exhibit structure (e.g. compositionality) (2005, pp. 391-392). In particular, he wants to show how one can account for these aspects of concepts without having to
posit abstract summary representations characteristic of traditional cognitive science (ibid., p. 394). Rather, he thinks that there is no need for explicitly stored summary representations if “diverse abstractions may be constructed online to represent a category temporarily” (ibid.).

Barsalou seems to think that if he can get away with not having to posit explicitly stored summary representations, he can avoid the traditional problems associated with trying to characterize such representations. But it’s not clear that Barsalou actually does have an account that avoids positing summary representations, for simulators seem to be summary representations. The question is how a simulator “establishes conceptual content”. Of course, once we accept the existence of simulators, the online representations which are constructed during particular simulations are dynamic and dispensable in the sense that they are tailored to one’s particular context and needn’t ever be used again. For example:

The entire content of a simulator is never activated at once; only a small subset becomes active on a given occasion…. Thus, on one occasion, the CAR simulator might produce a simulation of driving a car, whereas on others it might produce a simulation of fueling a car, of seeing a car drive by, and so forth. Although all the experienced content for a category resides implicitly in a simulator, only specific subsets are reenacted on a given occasion (2005, p. 400; cf. 1999 p. 586).

But the particular simulations on particular occasions only count as CAR simulations because they have been constructed from the CAR simulator, since it is the conceptual content of the CAR simulator that makes the particular simulations about cars (and not something else). Thus, the real work being done in Barsalou’s account is not the

33 The various troubles extant accounts of conceptual representation face is relative to the account. According to Barsalou, GOFAI theories face the problem of specifying the content of concepts (2005, p. 394) whereas connectionist theories have struggled with accounting for structured representations (ibid., p. 396).
particular simulations so much as it is the simulators from which those simulations are constructed (c.f. 1999, p. 588). The simulations may be dynamic and dispensable but the simulators themselves, it seems, cannot be. What seems critical for simulators is that they have definite boundaries for if they didn’t (i.e. if there was no one CAR simulator as distinct from a TRUCK simulator or a MOTORCYCLE simulator) then the particular simulations run would not be able to account for the type-token mapping that Barsalou wants to be able to account for. For example, Barsalou discusses how we might represent a thought like “an airplane has a nose.” The idea is that the NOSE simulator “interprets” a region of the object simulated by mapping a simulation of the NOSE simulator to a region of a simulation from the JET simulator (p. 404). But what makes the particular simulations count as instances of NOSE and JET is just that they have been simulated from the simulators that represent noses and jets. So the question is: how does the simulator come to represent objects such as jets and noses and relations such as above? Barsalou is not clear on the details of how this happens but in many places he talks about objects of the same category activating similar neural states (2005, pp. 400, 403; 1999, pp. 587). However, if similarity is supposed to establish discrete simulators, then it is insufficient to do so, as I will presently show.

As noted, a simulator is equivalent to a concept for Barsalou and although he doesn’t explicitly discuss the developmental aspect of simulators we can extrapolate from some of the things he does say. The account of how simulators are formed from percepts needs to offer a solution to the problem of abstraction: how, upon encountering instances of dogs, do I form the simulator DOG? There are some parallels between Mandler’s and

---

34 See figure at the end of this section, below.
Barsalou’s account of concept formation. In both theories there is the emphasis on the selective attending to some aspect of an object or event and a concept being created as a result. In Mandler’s case a conceptual representation (image schema) is abstracted from a percept whereas in Barsalou’s theory the percept is stored as is in a particular simulator. But selective attention isn’t enough to explain how a particular simulator is set up in the first place. So how do we get simulators?

Barsalou’s account of how simulators become established seems to be one according to which there is a process of accumulating percepts and storing them according to how similar they are to each other in terms of the pattern of features encoded by conjunctive neurons:

As multiple instances of the same concept are encountered, they tend to activate similar neural states in the feature maps...As a result, similar populations of conjunctive neurons tend to capture these states...Over time, conjunctive neurons integrate sensorimotor features across diverse category instances and across diverse settings, establishing a multimodal representations of a category (2005, p. 400).

There are two issues one might raise concerning the formation of simulators. First, how do the different modalities become integrated so that, e.g. either the feel of sitting in a chair or the look of a chair can both represent chairs? Surely the feature patterns captured by the conjunctive neurons will be very different for the look of the chair as opposed to the feel of sitting in a chair.\(^{35}\) But if these patterns are very different from one another, in virtue of what do they belong to the same simulator, CHAIR? Barsalou (1999, 2005) doesn’t attempt to answer how multimodal information could be integrated within the same simulator. Perhaps one could integrate them by “linking” both perceptions (the

---

\(^{35}\) This problem is a manifestation of the kind of binding problem I introduced in chapter one.
look and the feel of the chair) to the same object on some occasion. Second, and more fundamentally, how could similarity with respect to the kinds of features capturable by conjunctive neurons establish a simulator for kinds like CHAIR? This question may be regarded as more fundamental than the first, since if it turns out that not even stored percepts within the same modality can establish a simulator then neither can stored percepts from different modalities. It might be argued, however, that the different modalities are precisely what enable us to differentiate, e.g., a chair from a table. One of the kinds of information stored in the simulator could be event information such as *sitting* on a chair—versus *writing* on a table. In this way, the information from different modalities could help to distinguish the different simulators. However, in chapter one I argued that adding new aspects would not solve the problem of abstraction.

The times when Barsalou does say things about how simulators are established he seems to fall back on the idea that the members of various categories are similar enough to each other to establish that category (2005, pp. 400, 403; 1999, pp. 587). But although it may very well be that chairs “activate similar neural states in feature maps” it doesn’t follow that the conjunctive neurons that capture these patterns would in themselves have the ability to set up a distinct simulator such as CHAIR since, for example, there are probably many tables or beds or desks whose representations, captured by conjunctive neurons, are also similar. For example, a wooden chair is in many respects more similar to a wooden table than it is to an upholstered chair. Even if all the stored visual feature patterns of chairs did cluster together (which is not at all obvious), there’s still no reason to think that the conjunctive neurons capturing these feature patterns would have the resources to exclude other similar patterns. At which point would the conjunctive
neurons “decide” to draw boundaries within the mass of patterns they capture? There seem to be two problems here. One problem is that the kinds of features capturable in feature maps, as pointed out earlier, are things like EDGES and such general features will be insufficient to capture the kinds of differences that simulators are supposed to be able to capture. The second problem is the question of why we should think that discrete simulators (i.e. with boundaries that exclude certain neural patterns) would be formed from only the mechanisms of our perceptual systems.

As I suggested in chapter one, it seems that any empiricist account of concepts will face something like the following dialectic. The empiricist wants to show how we can form the concepts we do as a result of different kinds of experience. However, it seems that there must be at least some innate constraints built-in in order for us to be able to communicate. For example, it must be the case that we are innately drawn to attend to the same aspects of experience (if Mandler is right, these would be different aspects of movement and spatial relations). It also must be the case that I would group things together in a similar way that you would. If I had the tendency to group trees, cats and telephones together whereas you found such a grouping unintelligible, then we would have a difficult, if not impossible, time communicating. I would never discover what you meant by your use of a word that stood for the class of trees, cats and telephones. Such considerations would suggest that if we are to share categories, and these categories are formed based on the similarity of their members, then in order for us to share similar categories we must share a similarity space, that is, a metric according to which we judge certain kinds of features as similar to each other (e.g. red is more similar to orange than yellow). But what kinds of dimensions characterize such similarity judgments? Perhaps
it’s things such as color, shape, texture, etc. But similarity in such broad respects is almost certain not to pick out the categories we do have. For example, Barsalou (2005, p. 403) discusses a NOSE simulator which includes both noses from animals as well as the nose of a jet airplane. But why are such perceptually dissimilar things stored within the same simulator? And on what basis would, say, the tip of my finger or the end of my pencil be excluded from being stored in the “nose” simulator? If things are stored in simulators based on similarity in broad respects such as shape, then it seems like gerrymandered categories will be constructed. On the other hand, if we try to build in more precise dimensions according to which we group things as similar then we are coming dangerously close to presupposing concepts instead of accounting for them in terms of a similarity space. Thus, we can’t assume that the features to which we are perceptually sensitive are things like wings and legs. Rather they must be things like edges and color. Hence the empiricist either stays true to his empiricist principles but can’t attain types representing the classes of objects to which our words refer, or he gets the types but abandons his empiricism in the process. Barsalou’s account seems to grasp the former horn of the dilemma since the kinds of feature patterns which conjunctive neurons capture are things like edges, color, and direction of movement (2005, p. 398).

This line of thought raises a fundamental question: why should we think that storing perceptual symbols together according to some similarity metric would create discrete categories at all? We could grant that perceptual symbols are stored in a similarity space in such a way that certain features of objects are “closer together” in similarity space but deny that such a similarity grouping would be sufficient to create discrete categories. Rather, we would have a seamless similarity space, with certain
features (on a given dimension) shading off into other features like a color scale. And if there is just one seamless similarity space, then on what basis could we draw boundaries in order to establish discrete regions of space corresponding to categories or classes of objects? As noted, Barsalou is committed to saying that perceptual symbols underlie our judgments, so it is important that the content of his simulators correspond to the meanings of our words. Without simulators carving up our similarity space into discrete bits, there will be no concepts and thus no judgments.

Consider Barsalou’s (2005, p. 408) account of how a relational simulator ABOVE is formed. As a result of attending to the spatial relationships between two regions of space (occupied by some objects) we store relationships of being above within a similarity metric of vertical, horizontal and depth dimensions. Relationships must sufficiently match some criterion for the above relationship, and a number of different combinations of vertical, horizontal and depth dimensions will be counted as ABOVE, which means they are similar enough to a simulation that can be constructed from the ABOVE simulator. But why should we believe that attending to these relationships stores them in a discrete simulator rather than in some more continuous similarity space according to more general dimensions (e.g. “spatial relationship”)? That is, ABOVE would seem to shade off seamlessly into other relationships such as NEXT TO and BEHIND, in which case we would just have a seamless 3-dimensional similarity space which would, perhaps, enable us to categorize some spatial relationships as more similar to others but would not enable judgments (such as X IS ABOVE Y) since there would be no discrete concepts (such as ABOVE) carving up the similarity space. If the principle by which simulators are created is that things are similar in respect R, then similarity,
being a graded notion, would seem to be insufficient to ground the formation of discrete categories. Barsalou continually remarks that multiple instances of the same category, spatial relation or property are stored together, which establishes a simulator (2005, p. 408). Presumably, they are stored together because they are similar enough in some respects (cf. 1999, p. 587). But if similarity is supposed to be the means of establishing a simulator, then it is insufficient to do so.

![Barsalou’s NOSE simulator representing “the jet has a nose”](image)

**An explanans sans abstraction**

The datum we are trying to account for is the fact that infants are able to distinguish global categories such as vehicle, animal and container. Suppose we accept Mandler’s account that prioritizes how things move (e.g., self movement) and spatial relationships
(e.g., containment) as the basis of an infant’s first concepts. Then what is needed is simply for the infants’ attention to be grabbed by different kinds of motion events and spatial relationships, to which they selectively attend. As a result, a portion (corresponding to the motion aspect of an event, captured by neurons that respond to motion information) of the overall pattern of neural activation is isolated and stored in long-term memory. This is a perceptual symbol; its virtue is that it can fulfill the role played by Mandler’s image-schemas (i.e. encoding only one small aspect of the overall perceptual array) without our having to postulate a separate, conceptual representational system. At this point, however, I am leaving Barsalou’s account since his perceptual symbols are always stored in simulators. Rather, I am supposing that the perceptual symbols are stored in a similarity space which has numerous dimensions corresponding to the features that our perceptual systems can isolate.

Consider a concrete example. As a result of being drawn to different kinds of motion events, infants will selectively attend to and store information about how animals move. These particular characteristics of animal movement are stored along different dimensions in the similarity space. The similarity space would contain dimensions for the kinds of movement/spatial information that Mandler suggests but it would be stored in the form of perceptual symbols rather than image-schemas. Recall that perceptual symbols are just stored representations of instances (i.e. a portion of the total neural activation that arises during perception) while image-schemas are representations of kinds (e.g. kinds of movement such as SELF-MOVEMENT and ANIMATE MOVEMENT). Thus what is stored in the similarity space is an instance not a kind. For example, an infant watching a cat will be drawn to various aspects of that cat’s movement
such as that it moves by itself (without anything acting on it), it moves in an irregular, unpredictable way, and interacts contingently with other things (e.g. with the infant itself). Perceptual symbols along the dimensions to which the infant selectively attends will be stored in the similarity space. Now consider what happens when the infant sees a dog. Since the aspects of the dog’s movement (i.e. moves by itself, moves in an irregular, unpredictable way, and interacts contingently with other things) will be similar to those of the cat, the perceptual symbols of the dog’s movement will be stored in the similarity space somewhere “near” the perceptual symbol created by attending to the cat’s movement (rather than one created by attending to a vehicle or plant).

It is important to emphasize that the perceptual symbol would contain only those aspects of the total perceptual state which are attended to; so if some features like color and texture are not attended to then there will not be corresponding features contained within the perceptual symbol. Likewise for the idea of containment: if what infants selectively attend to are only certain aspects of the teacup used to give someone a drink, then we should not be surprised if they are just as likely to imitate this event using a frying pan as using a teacup. On Barsalou’s picture, a portion of “qualitative neurons” which respond to spatial relations such as containment (which are selectively attended to by the infant) is stored as a perceptual symbol. The perceptual symbol (the neural “fingerprint”) of the teacup and frying pan will be similar since, \textit{ex hypothesi}, infant’s attention is drawn to spatial relations and the teacup and frying pan are both containers. Again, perceptual symbols differ from image-schemas in that they represent instances, not kinds. Of course, Barsalou countenances simulators as well, which I have claimed play the role of kind representations for him, but I do not countenance kinds and have
substituted instead the idea of a similarity space. The representations in the similarity space are identified not by their content (as image-schemas are) but by their location in the similarity space.

Finally, how would we describe what Mandler’s infants are doing in the object-examination paradigm? Why do they distinguish between animals and vehicles but not dogs and fish? On this view, the increased handling times result from the fact that a certain object is seen as novel, but an object is seen as novel not because it doesn’t fit into a prelinguistic concept that the other objects do fit; rather, it is seen as less similar to the other objects with respect to certain dimensions that are salient to the infants. However, since the animal figurines are not moving in the experiments, infants must rely on featural information other than how the things move to categorize the models as animals or vehicles. For example, perhaps the presence of self-movement becomes associated with the presence of moving parts (i.e. wings, legs or fins) and it is by these features that infants in the experiments categorize the models of animals as similar with respect to moving parts. Consider the object-examination test where the infants dishabituate to an airplane when they were habituated with a bird (but not to a fish when habituated with a dog). The salient features for infants are things such as wings and legs—those things that are associated with animal movement, self-movement and contingent interaction. So, for example, a dog is more similar to the bird than to the airplane with respect to presence of movable parts (which are, however, not currently moving). The airplane is seen as novel since what infants selectively attend to when animals and vehicles are compared are the things like the presence of movable parts. But the fish, unlike the airplane, has movable
parts like the dog. Thus, according to the paradigm, the infant would be less likely to handle the fish since she doesn’t find it novel.

David Rakison (2005) has argued, contra Mandler, that the “movement information” which Mandler thinks supports the view that infants have two different kinds of representational systems, needn’t be understood as entailing a fundamentally different system. He has carried out versions of Mandler’s generalized imitation paradigm that suggest that what accounts for children’s behavior in the paradigms is not some abstract “image schema” which summarizes types of motion but, rather, the presence of movable parts. His idea, roughly, is that knowledge of motion properties is the result of associations between parts and types of movement with which those parts become associated. But at first, before these associations have been formed, children rely mainly on the presence of certain parts such as legs or wheels. For example, Rakison created hybrids of animals and vehicles (e.g. animals with wheels and vehicles with legs) in order to investigate children’s categorization preferences. The exemplars from which the children could choose to imitate the modeled event were of four types (relative to the object used to model the event): same part/same category, same part/different category, different part/different category, and different part/same category. Rakison found that whereas 18 month old infants used both SPSC and SPDC, 22 month olds used exclusively SPSC. Thus, 18 month olds were willing to transgress the animate/inanimate distinction yet always picked exemplars with same parts and, conversely, would not pick exemplars with different parts even if they were of the same category. Thus, for example, infants would generalize from a dog to a cat but not a dog to a dolphin. This suggests that infants initially categorize according to presence of parts rather than some
more abstract “animal” concept based on movement information. Only later (22 months) did their categorization respect the animate/inanimate distinction.

I offer Rakison’s empirical evidence in support of my above account, according to which infants are categorizing based on the presence of certain parts. Rakison’s research suggests that the similarities according to which infants categorize may just be the presence of certain kinds of moveable parts, which later in development come to be associated with certain kinds of movement. Nevertheless, inasmuch as Rakison thinks that his one representational system can provide a mechanism for conceptual development, his theory too would be susceptible to the same problems as Barsalou’s.36

In this chapter I have attempted to set out a view according to which categorization need not entail conceptualization. To have a concept is to be able to make a judgment containing that concept and one can have the ability to categorize without thereby having the ability to make judgments. This, I suspect, is just the position that Mandler’s infants are in: they can categorize but they can’t make judgments like “x is an animal.” I have claimed that one can understand the categorizations that prelinguistic infants are making without having to posit kind representations such as ANIMAL or CONTAINER. Rather, instances are represented in a similarity space according to dimensions which are plausibly isolated by mechanisms of our visual system such spatial relationships. A similarity space cannot represent a kind because the dimensions which characterize it (i.e. those dimensions which are plausibly isolated by our visual and other sensory systems, such as color, shape, and spatial relationships—and not doors, wings or

36 Rakison thinks that a domain general associative mechanism forms “categories that cohere because of similarity relations” (p. 137, c.f. Rakison, 2003). Thus it seems his view would indeed inherit the same problems of Barsalou’s.
noses) are too broad to be able to carve any joints that would correspond to concepts like ANIMAL or VEHICLE. Importantly, there is no principle for drawing boundaries within areas of similarity space, and we need boundaries for conceptual representation. Nevertheless, instances stored within a similarity space can be more or less similar to one another depending on their particular neural fingerprint.

But although I reject both Barsalou’s and Mandler’s kind representations (simulators and image-schemas), I accept Mandler’s idea that the prelinguistic infant’s thinking about objects may be disproportionately biased towards spatial and movement information rather than color and shape information. And in building my own account of how infants might understand objects I have drawn on Barsalou’s idea of storing portions of the total neural activation, where the portion stored is the aspect attended to.

Mandler’s idea also fits in here since she proposes that infants consciously attend to the spatial and movement aspects of objects.

My account of how the infant might be representing objects in thought in terms of a similarity space also serves as a substantive suggestion of what nonconceptual thought might be. The idea is that thinking is a matter of comparing instances within a similarity space and this contrasts with thinking as forming judgments.
CHAPTER 3: THE OBJECT CONCEPT

There is a certain picture of cognitive development (in particular, the development of sortal concepts) one can find in recent developmental literature, especially in the work of Fei Xu (1997, 2002, 2005), which goes something like this: infants start out with an initial “object concept” which is an innate or early developing understanding of objects of a certain kind, namely Spelke-objects (bounded, coherent, three-dimensional physical objects that move as a whole). This (supposedly conceptual) understanding is attributed based on children’s expectations about how objects are individuated. “Individuation” is a term of art, which essentially describes the infants’ longer looking times at unexpected outcomes in the violation of expectation paradigm (e.g., when only one object is present when two objects would be expected). Intuitively, individuation is the ability to recognize two objects as distinct. This initial object concept is characterized in terms of spatio-temporal information (as explicated below).

As they develop, infants begin to make differentiations within this initially undifferentiated object concept. Although at first infants do not use property information (i.e. shape, color, texture) or kind information (i.e. duck, ball, cup) as evidence for distinct objects, they later are able to individuate objects based on such information. Fei Xu 2005 thinks that there are two different types of systems for individuating objects: an object-based system, which uses only spatio-temporal information, and a kind-based system, which uses kind representations. The object-based system functions essentially as an initial ordering and taming of the blooming, buzzing confusion into discrete bits. It is because children are able to initially represent individual objects (Spelke objects) that
they later come to represent objects as belonging to particular kinds. Xu 2005 thinks, furthermore, that language has a particular role to play in getting infants from the initial object concept to kind concepts. In particular, words function as an “essence placeholders” so that infants look for similarities between objects which are called the same thing.

Is it correct to speak of the spatio-temporal individuation skills of infants as a conceptual ability? Carey and Xu 2001 have claimed that these object tracking skills should be viewed as more conceptual in nature, while others have denied it (Bernal, 2005; Scholl and Leslie, 1999a). I will claim that the infants’ so-called object concept, is not a concept at all because it doesn’t fit the characterization of conceptual thought I laid out in chapter 1. That is, the ability to individuate objects does not imply judgments being made.

Object individuation in infancy

Scholl and Leslie note that there are two different aspects to the child’s so-called object concept: the aspect which describes their expectations about contact mechanics and the

\[\text{\textsuperscript{37}}\text{C.f. Spelke 1990: “a partial segmentation of surfaces into objects might ease the task of object recognition by limiting the potential matches of object models to visual arrays. …Consider the example of the horse and rider. Although the principles of cohesion, boundedness, and no action at a distance to not specify where a horse ends and its rider begins, they limit the possible boundaries of those objects. No object model could correspond to the horse’s moving leg and the stationary ground beneath it, or to the rider’s head and a tree behind it” (p. 50).}

\[\text{\textsuperscript{38}}\text{What Xu actually says is the converse of this: that objects called by different names will have different properties. But it seems that in order to establish differences one must already have some sort of criterion or metric according to which properties of objects are similar.} \]
aspect which describes their expectations about spatio-temporal constraints (1999a, pp. 28-29). On the one hand, infants have expectations about how objects should interact when they contact. For example, infants seem to expect that solid objects cannot pass through each other, as evidenced by their longer looking times in cases where one solid object appears to pass through another solid object. For example, Baillargeon et al. 1985 habituated five month old infants to a screen which moved in a 180 degree arc. Then they placed an object behind the screen, such that the screen would not be able to make the full 180-degree rotation. In this condition, the infants looked longer when the screen made the full 180-degree rotation than when it stopped short of making a full rotation (as would be expected since the object would keep it from making the full rotation). On the other hand, infants seem to have expectations about spatiotemporal constraints on objects, such as that objects must trace a continuous spatio-temporal path. The notion of an infant’s “object concept” is supposed to encompass both contact mechanics expectations and spatio-temporal expectations. The ability to individuate objects, however, is primarily an ability that describes expectations based on spatio-temporal information, and it is with this aspect of the child’s understanding of objects that the present chapter mainly deals. Later in the chapter I will return to the contact mechanics aspect in examining the argument of Carey and Xu 2001 that the child’s object concept is indeed a concept.

Some of the earliest work in object individuation was carried out by Elizabeth Spelke. Spelke 1990 discusses a number of different results in which she used occlusion displays to test whether 3-5 month old infants used property information or movement information in order to segment or individuate objects. For example, imagine what looks
to be a rod, which is occluded by a screen so that only the ends of the rod are visible. Now the two ends of the rod moves behind the screen together and then the screen is removed to reveal either a connect rod or two unconnected rods. In such a case, infants look reliably longer at the unconnected rod than at the connected one. However, if the occluded object isn’t moving, then infants looked equally long at the connected object and the unconnected object (once the occluder was removed). Importantly, as long as the object was moving, the same looking time response was found even if the two ends of the object were perceptually dissimilar:

Any unitary translation of the object in three-dimensional space led infants to perceive a continuous object: Vertical translation and translation in depth had the same effect as lateral translation. Perception of a moving, center-occluded object was not affected by the object’s configurational properties: Infants perceived a connected object just as strongly when the object’s visible surfaces were asymmetric and heterogeneous in texture and color as when they formed a simple shape of a uniform texture and color (Spelke, 1990, p. 33).

Conversely, in studies investigating the infant’s perception of boundaries, infants expected objects to be separate when they moved independently of one another, further supporting the idea that object segmentation/individuation in younger infants is due to spatiotemporal information only (Spelke, 1990, p. 36). Fei Xu and Susan Carey have developed new paradigms for studying infants’ object individuation abilities. The following two experiments are foundational in Xu’s development of her theory of cognitive development.

**Experiment 1**

Consider the following experiment. A four and a half month-old infant is presented with two screens on a small stage (there is a space between the two screens). A duck emerges
from behind the left screen and then returns behind it, then an identical duck emerges from behind the right screen then returns behind the screen. This process takes place a few more times before the screens are lowered, revealing either one of two outcomes: a duck behind only one of the screens or a duck behind both screens. Adults expect there to be two ducks, since it would not be physically possible for the same duck to emerge from behind each screen (since they are separated by a space, and if the duck were to traverse that space between the two screens, one would see it). It is therefore surprising that there should be only one duck when the screens are lowered. In the preferential looking method used in these experiments, the dependent variable used to measure infants’ responses is the amount of time that the infants look at either the expected outcome or the unexpected outcome. Results across multiple replications have given a robust result: infants always look longer at the unexpected outcome of only one duck than at the expected outcome of two ducks.\(^{39}\) From this experiment one can see that infants are able to individuate objects based on spatiotemporal information (the spatiotemporal information relevant here is that no object traversed the path between the two screens).\(^{40}\) But such a result is consistent with the supposition that infants were tracking the ducks either as ducks or as a particular configuration of properties. The second experiment is supposed to rule out this possibility.

\textit{Experiment 2}


\(^{40}\) A variant of this spatiotemporal condition uses only one screen but shows both objects simultaneously, thus giving unambiguous spatiotemporal information that there are two objects. In this variant, infants perform the same as they do in the two screens version explained here.
Consider the following experiment, which uses the same violation of expectation-preferential looking method. Ten month-old infants view one large screen (as opposed to two smaller ones as in the previous experiment) on the stage. Infants then watch a duck appear from the right side of the screen and then disappear back behind the screen. This is followed by a ball appearing from behind the left side of the screen and then disappearing back behind the screen. As before, these steps are repeated and then the screen is removed, revealing one of two outcomes: either there is a ball and a duck or there is just a duck. In this case, the adults would expect there to be both a duck and a ball behind the screen. However, infants do not look any longer at the duck-only outcome than at the duck and ball outcome. However, if infants are shown both the duck and the ball emerging from separate sides of the screen simultaneously, they will look at the duck-only outcome longer. These results suggest that for 10 month old infants, spatiotemporal information alone is used in individuation, since property information of there being two objects was given, but no spatiotemporal information was given.

The interesting finding here is that infants do not take the differing property information as evidence for distinct objects. Xu and Carey 1996 claim that it is not the case that the infants didn’t notice the property differences between the duck and the ball, they just failed to use these property difference as evidence for distinct objects. They claim that infants recognized the property differences since it took longer for them to habituate to a duck and a car appearing repeatedly from behind the screen than it took for them to habituate to just a duck appearing alternately from different sides of the screen. (Habituation is achieved when the infant loses interest in the display.) Xu 2005 claims that essentially what is happening here in the case of these ten month old infants is that
the object-based system (which uses spatio-temporal information alone) overrides the property information (such as differences in color and shape).

However, subsequent research has been equivocal with regard to the claim that ten month-old infants cannot use property information (i.e. color, shape, texture) to recognize two objects as distinct. In particular, this claim has been called into question by Renee Baillargeon and her colleagues (Wilcox and Baillargeon, 1998; Needham and Baillargeon, 2000). The gist of the Baillargeon studies is that if the task is made simpler, then ten month-old infants (and even younger) can succeed in individuating objects based on property information. Success at individuating objects using property information, for Baillargeon and colleagues, is not due to the child’s acquiring kind concepts (as Xu and Carey have claimed) but as a result of increased information processing capacities. That this is so is shown by the fact that even very young infants can use property information to individuate objects so long as the task is less taxing on informational processing systems.

For example, Needham 1998 used looking time to determine whether 4.5 to 6.5 month old infants, who were familiarized with an adjacent display of two objects with differing textures and colors (ambiguous as to whether they were connected), expected the objects to move together or move apart when an experimenter dragged one of the objects. They found that infants looked longer at the “move-together” trial, suggesting that infants as young as 4.5 month can succeed in object individuation based on property information. Wilcox and Baillargeon 1998 introduced a distinction between event mapping and event monitoring to explain why infants might succeed in using property information to individuate objects in tasks such as Needham (1998) but fail to in tasks
such as Xu and Carey (1996). The basic idea behind the distinction is that there is more strain on an infant’s information processing capacities when going from an occlusion to non-occlusion event (event mapping), as in the Xu and Carey studies, than when only dealing with a non-occlusion event (event monitoring), as in the Needham (1998) study. The claim is that event mapping tasks are more taxing on the infant’s informational processing capacities and that this is what accounts for the different findings. Further supporting the informational processing hypothesis, Wilcox and Baillargeon 1998 showed that the trajectory of the object also affects the difficulty of the task. For example, they showed that 9 month olds can succeed in the Xu and Carey paradigm, so long as the path of the object is “simplified” (which means just that there are no reversals of the objects appearing from behind the screen). In the multiple reversals version, however, 9 month olds failed to individuate, whereas 11.5 month olds were successful (just as in Xu and Carey 1996).

Xu 2005 has her own way of accounting for these different findings. As noted, Wilcox and Baillargeon brought the objects out only once whereas Xu and Carey brought the objects out multiple times (4-7 times). Xu 2005 suggests that the multiple times created stronger statiotemporal bias that there was only one object and that this stronger evidence overrides the property information in the Xu and Carey study. One can create similar illusions in adults, such as making it look like the red square turned into a green circle if the green circle emerged from behind an occluder following the same path as the red square was. Similarly, Xu’s idea is that multiple repetitions made it seem as if there was only one object which was changing properties, in contrast with the single repetition method which didn’t create this statiotemporal bias and so the property information was
not overridden. I don’t think this explanation settles the matter, the main problem being that we don’t have any independent evidence for Xu’s claim that multiple repetitions would establish stronger spatiotemporal evidence than single repetitions (i.e. it seems ad hoc).

Nevertheless, regardless of how the debate about the age at which infants can use property information turns out, Xu still has another question which needs an answer: Why do 12 month-old infants succeed at the very same paradigm in which 10 month-olds fail? As noted above, Baillargeon and her colleagues think that the difference can be explained in terms of the more limited information processing capacities of younger infants. Xu, on the other hand, thinks that the 12 month-olds must be using kind information (i.e. kind concepts) whereas 10 month-olds do not. Her argument for this can be cast as follows:

1) If infants are not using spatiotemporal information for individuation then they must be using either property or kind information
2) They do not use spatiotemporal information
3) So they must be using either property or kind information
4) But they don’t use property information
5) Therefore, they use kind information.

The main evidence she gives for premise 2 is a series of studies in which 12 month-olds failed to individuate objects by using only property information. For example, using the

41 Xu’s talk of “using property information” or “using kind information” is potentially misleading. I take it that this talk is a gloss of something like, “representing the object in terms of its various properties (color, shape, texture, etc.)” or “representing the object as a kind (e.g. DUCK). Kinds, for Xu’s purposes here, are sortal kinds, like DUCK, CUP and BALL.
one screen paradigm outlined above, two different colored balls were used as the objects and in this condition, 12 month olds did not look any longer at the unexpected outcome, suggesting that the property differences (different colors) were not sufficient to make them expect two objects rather than one. They replicated this finding with different shapes within a category (i.e. a sippy cup with two handles vs. a cup with one handle—both from the CUP category). However, when two different shapes from a different category were used (regular cup vs. a bottle), then the infants did look longer at the unexpected outcome. Xu takes this evidence to show that the infants are not using property information to individuate objects but, rather, kind information. But it seems to me there is an alternate explanation in the offing for someone who would maintain that it is, in fact, the property differences that are responsible for the infants’ individuation behavior (and not kind concepts). The explanation would be that there is a kind of threshold of difference that must be crossed for the infant to recognize the objects as distinct. One of the characteristics of crossing kind boundaries is, after all, often times also a matter of larger differences in perceptual similarity. Thus, a red ball is more similar with respect to shape and texture, etc. to a green ball than it is to a rubber duck. So why not just say that in the duck vs. ball case there were more (and more salient) differences than with the different colored balls case?

I can’t help but notice a similarity here with recent literature on change blindness. Some of the results of change blindness studies are remarkable: people not noticing a man in a gorilla suit walking through a scene, people not noticing the surreptitious switch

\[\text{Xu, Carey, Quint (2004).}\]
of the person they are talking to, etc. Some of these findings suggest that people notice far less than they think they do and that there are good reasons for this (e.g. the decrease of information processing load). Amazingly, one can switch the person the subject is talking with in these studies without the subject noticing, but I suspect you could not just change anything you wanted without the person noticing. For example, it seems you could not switch a 4-foot tall person with a 7-foot tall person and have the subject not notice. Nevertheless, that the subject noticed the change here would not entail that they were using concepts such as TALL PERSON and SHORT PERSON to notice this difference. (Of course, they might later describe the difference in this way, but that doesn’t mean their initial noticing of the difference was dependent on such a description.)

Likewise, I want to say, the fact that the infants noticed the difference between the duck and ball (but not between the different colored balls) does not show that they used a kind concept to do it. One can imagine an infant study (analogous to the change blindness study I’ve imagined) in which, instead of different colored balls, drastically different sized balls are used, one very small and one very large. I suspect that such drastic differences in size would be sufficient for infants to individuate the objects. And, here, I think, we would be less inclined to say that successful individuation was due to the infant having the concepts LARGE BALL or SMALL BALL. But, if so, this would show that property differences can cause an infant to individuate objects as long as the differences pass some threshold of salience.

---

44 Xu, Carey and Quint 2004 did use different sized balls as well and found that the infants didn’t individuate them. But I’m imagining pushing one aspect (size) to an extreme analogous to switching a 4-foot tall person for a 7-foot tall person in the change blindness study. Perhaps the child wouldn’t notice small difference in size but it seems likely they would notice large differences.
Thus, it seems one can account for the individuation behavior of infants in terms of salient property differences, in which case premise 2 does not hold and the argument that infants must be using kind concepts fails. There remains the question of which differences are salient and where the threshold is, but that is an empirical question that is in need of further research.\(^{45}\) My present point is just that Xu can’t get away with claiming, based on the evidence she cites, that infants could not be using property information in the individuation tasks.\(^{46}\)

**Is the “object concept” a concept?**

My current task, however, is to explore how best to think about the spatiotemporal individuation ability that infants display. It is clear that the infants in these experiments are singling out an object and are paying attention to it (i.e. their heads follow the objects moving across the stage). Moreover, they have certain expectations for particular

\(^{45}\) Some research bearing on this question already exists and would seem to support my contention that there are salience thresholds. For example, Wilcox 1999 showed that, in fact, infants are able to use size and shape to individuate objects much earlier (4.5 months) than they are able to use patterns (7.5 months) or colors (11.5 months). Of particular interest for assessing Xu’s claims is that infants’ ability to use color alone to individuate objects came much later than the ability to use shape or size, which suggests that at younger ages, shape and size are more salient than colors. Tremoulet, Leslie and Hall 2000 found similar results: 12 month-old infants noticed shaped changes but not color changes.\(^{46}\) Indeed, Xu 2005 admits this herself, which makes her argument even more puzzling. She will say that her claim applies to the paradigm which uses multiple repetitions of the objects moving out from behind the screen and then back (since, as noted above, she admits that infants can use property information to individuate in the Wilcox/Baillargeon single repetition paradigm). But why couldn’t more salient and stronger property differences override the spatiotemporal bias in Xu’s multiple repetition studies? I have argued above for an interpretation of the data that is consistent with this hypothesis. Xu would have to argue for the claim that property information cannot override the spatiotemporal bias in the multiple repetition paradigm but that it can in the single repetition paradigm. I see no prima facie reason to draw an in principle distinction between the use of property information in the two paradigms.
outcomes, which, when violated, cause the infants to look longer. However, as experiment 2 shows, they sometimes do not use property or kind information, but only spatiotemporal information, in individuating objects. The idea that infants sometimes individuate objects by spatiotemporal information instead of property information is further supported by Elizabeth Spelke’s seminal studies.

**Carey and Xu’s arguments that the object concept is conceptual**

Is there, then, an object concept, SPELKE-OBJECT, which is characterized in terms of spatiotemporal properties it represents? Carey and Xu 2001 (c.f. Xu, 1997) argue that the child’s object concept is indeed conceptual and “thought-like.” They claim that the child’s object concept, like any other concept, has both wide content (extension) and narrow content (conceptual role): wide content because there is a certain class of objects which infants track, viz. the extension of Spelke-objects, and narrow content because the object concept plays a role in articulating physical reasoning and intentional action (2001, p. 206).

That Spelke-objects have wide content is shown by the fact that infants don’t track just any kind of objects. Consider the following experiment: eight month-old infants viewed sand being poured onto the ground and then hidden by a screen. Then another, separate, screen was introduced and another pile of sand was poured behind it. The screens were then removed revealing one of two outcomes: one pile of sand (unexpected outcome) or two piles of sand (expected outcome). Eight month-old infants did not look any longer at the unexpected outcome than at the expected outcome. However, when sand-pile-objects (perceptually identical to the piles of sand) were
lowered behind the screens, infants succeeded in looking longer at the unexpected outcome (cited in Carey and Xu (2001) p. 207). Apparently, eight month-old infants do not track certain objects such as piles of sand being poured. It is not clear exactly what the extension of this so-called object concept is, but whatever it is, it will be discovered empirically. The class of Spelke-objects is just whatever infants would track in the aforementioned paradigms. Of course, wide content is not sufficient to establish something as conceptual content, since, as Bernal 2005 notes, informational (wide) content can be carried by many non-conceptual things, including thermometers. In general, that something tracks what we recognize as a certain class of things, doesn’t mean the thing has concepts. Thus, if the child’s object concept is really a concept, then it must be because of its narrow content.

Carey and Xu 2001 give a couple reasons to think that the object concept is more thought-like. Consider the following paradigm (Xu and Carey (2000) p. 290-291; Carey and Xu (2001) p. 189). An experimenter reaches through a slit in a box, whose contents the ten month-old infants cannot see, and pulls out a toy telephone. The experimenter puts the telephone back into the box and then pulls out a toy duck, and then places it back into the box. The infant then reaches into the box only once and pulls out either the telephone or the duck. In contrast, when both objects are shown at the same time, the infants will reach into the box twice (and no more). In addition, when the infants are

---

47 Whatever the extension is, however, it seems clear that the extension of the child’s concept SPELKE-OBJECT will differ from an adult’s concept OBJECT. Fodor (unpublished draft) notes that many objects do not count as Spelke objects simply because they do not move (e.g. trees, buildings, large rocks, a brick which is now part of a wall), as does Wiggins (1997, p. 415). This highlights a difference between the extensions of the infant’s concept SPELKE-OBJECT and an adult’s concept OBJECT. A building would count as an object for an adult but not as a Spelke-object for an infant, since it doesn’t move.
shown only one object, they will reach *only once* into the box. These results support the other findings of the violation of expectancy method, but also show something more: the child’s representation of the *unseen* object(s) can guide her volitional action, and Carey and Xu take the guiding of volitional action as something characteristically thought-like.

Consider another experiment. Feigenson, Carey & Hauser, 2002 had ten month-old infants watch as graham crackers are placed into a box (into which they cannot see). The question is whether they would crawl to the box containing the larger number of graham crackers, which they did, but only when the choice was between 1 vs. 2 or 2 vs. 3 crackers. Infants consistently crawled to the box containing more crackers when the choice was between 1 vs. 2 and 3 vs. 4 crackers, but not when the choice was 3 vs. 4 or 3 vs. 6. Here, again, the child seems to represent the unseen objects and this representation guides her action. The abilities displayed by the 10 month-old infants in this study are conceptual in nature, Carey and Xu argue, because they involve holding onto representations *while the corresponding objects are out of sight* and manipulating them such that those manipulations guide volitional action.

The interesting capability described in these two studies is the ability to hold in memory some sort of representation of the objects such that this representation affects the infant’s subsequent behavior. For example, how are we to explain why the infant reaches into the box twice and no more, when unambiguous spatiotemporal information of there being two objects has been given? How are supposed to explain behavior which is so nicely adapted to the environment? Well, the idea is supposed to be that the child reaches no fewer or no more times into the box than there are objects in the box because the child
represents how many objects are in the box.\textsuperscript{48} The question that needs to be asked, however, is what is the nature of this representation? If the child were representing the objects conceptually, then what kind of thought might she be having? Perhaps the infant would be using SPELKE OBJECT to represent each of the objects used in the experiments, such as in the thought “there are two Spelke-objects in there.” But if this were so then we’d have to explain where other concepts such as TWO and IN came from. Carey and Xu’s own idea seems to be something like that the infants have a representation of Spelke objects which function as “short term memory models” which (in the graham crackers experiment) they use to compute which box contains more crackers—e.g. compare the two Spelke objects in the one box with the three Spelke objects in another box (2001, p. 205).

But for this kind of picture of what is going on it does not seem that one needs to view these representations as conceptual. Suppose, rather, that what the child is doing is comparing percepts, stored images. Indeed, this would make more sense of the fact that infants are only able to represent the difference when the contrast includes smaller numbers (1-3), since in this case we can explain the difference as a kind of subitizing ability—an ability which only works with smaller numbers (Kaufman, Lord, Reese, and Volkmann, 1949).\textsuperscript{49} Moreover, as Sara Bernal notes, the fact that a representation can guide volitional action does not thereby make it conceptual, as can be seen by the fact that representations which are characteristically perceptual can also guide action. For

\textsuperscript{48} Although, as the graham crackers experiment shows, infants only succeed with smaller numbers of objects.

\textsuperscript{49} The kind of subitizing ability I am envisioning would be temporally extended, i.e. the hand reaching into the box not represented as a discrete number of times but, rather, as a kind of temporally extended magnitude.
example, imagine being shown two different color tiles and then being asked, once both of them had been taken away, which was brighter. This would seem to be a clear case of one’s action (the answer to the question) being guided by comparing images or percepts in memory (Bernal 2005, p. 304). Thus it seems that the ability to guide volitional action is not uniquely characteristic of conceptual representation in general, and in these particular cases it does not seem that invoking conceptual representation is the best way to explain the infants’ behavior, since explaining the thought in which SPELKE OBJECT is used would presumably require positing other concepts as well.

So much for the guiding of volitional action being a characteristic of concepts. There remains Xu and Carey’s claim that the object concept “articulates physical reasoning.” This claim is related to the other aspect of the child’s “object concept” which I noted at the beginning of this chapter, namely, that which describes the child’s expectations regarding contact mechanics. The idea here is supposed to be that infants have a rich set of expectations about the behaviors of objects, which go beyond the mere tracking of objects: expectations about the solidity of objects, expectations about how objects will behave on contact, and cross modal expectations (Streri and Spelke, 1988).\(^50\) Furthermore, these expectations are supposed to derive from a single representational mechanism, whose diverse effects lead us to view it as more conceptual in nature. The contrast here, seems to be with simpler, single purpose representation, such as edge detectors, and the point is supposed to be that the object concept plays many different roles and so cannot be assimilated to a simple, single purpose representation. The point, I

\(^{50}\) Children were tested using the same violation of expectation paradigm, except that the habituation was haptic and the test trials were visual. Infants were surprised (looked longer) on the unexpected outcomes of the test trials—i.e. where the haptic and visual properties of the “same” object did not match up.
think, is not merely the expectations, but the diversity of expectations which a *single* representational capacity is supposed to explain.

One problem with this argument is that we have not been given a reason to prefer the view of the object concept as a single representation to the view of a network of simpler representations. As Bernal notes, one could just as easily explain the infant’s object perception using a network of simpler mechanisms as opposed to one central representation which explains the diversity of effects:

How might we account for [capacities not accounted for by a simple visual indexing mechanism], if not with the single powerful mechanism that Carey and Xu envision? One methodology would be to seek to explain each of these subcapacities with the least thought-like mechanism that is adequate to the task. On the picture that would result, object perception would be underwritten by a patchwork of ‘dumb’, narrowly specialized mechanisms, each one fitting the same generally modular profile of Indexing: a confederacy of dunces (2005, p. 304).

If such a case is possible and, perhaps, more evolutionarily plausible, then the argument from the complexity and diversity of various kinds of expectations falls apart since these need not be due to one mental representation mechanism, but to many simpler ones. For example, some of the individuation experiments involve recognizing that objects which move together (when hidden by an occluder) are parts of the same object (Spelke, 1990). We might think of this as a kind of Gestalt principle for the perception of objects based on spatiotemporal information. Other individuation experiments (i.e. the graham crackers experiment) involve, so I have suggested, a kind of subitizing ability. But these two abilities could easily be carried out by separate mechanisms, in which case we would have a “confederacy of dunces” explanation rather than positing one more centralized representation, which is responsible for both abilities.
Carey and Xu give a further reason for thinking that infants’ knowledge of Spelke objects must be conceptual: infants learn new generalizations about the behaviors of objects. For example, infants do not initially expect that unsupported objects will fall. At first, they expect only that there must be some contact from below, then that at least half of the object’s base must be supported, and finally they take into account the geometry of the object. But now consider how infants can arrive at such general knowledge. In order for infants to learn new things about the behaviors of objects, one might think that they must have the ability to generalize over a certain kind of thing, namely a Spelke object. It seems that Carey and Xu’s argument is something like this: knowledge about objects is stored in the form of a generalization, which requires a predicate to be involved in the general thought, so there must be such a thing and SPELKE-OBJECT fits the bill. “Physical reasoning about Spelke-objects embodies knowledge formulated over the category object, whatever the format of this knowledge” (2001, p. 209). Of course, if this picture is taken literally, and the infant’s knowledge is to be characterized as a generalization, then infants must also have concepts such as UNSUPPORTED and GEOMETRIC CENTER, in addition to the concept SPELKE-OBJECT, since they would also be contained in the generalizations. However, although some have claimed that learning must be a process of forming generalizations by hypothesis formation and confirmation (e.g., Fodor 1975), there are other ways to account for learning without having to couch it in the talk of generalizations. Although it may be that some kind of ability to segment an array into objects is required for one to learn about them, it does not seem that this individuation ability need be a conceptual representation SPELKE-

OBJECT, unless one already accepts a model of learning according to which pieces of knowledge are generalizations using concepts. So the argument that infants’ learning about objects requires us to admit that there is a concept OBJECT (or SPELKE-OBJECT), depends on a further claim about the nature of learning and knowledge.

The object concept measured against some characteristics of conceptual thought

The upshot of my criticism of Carey and Xu 2001 is that some of their criteria seem not to be good criteria of what counts as a concept. In addition to criticizing Carey and Xu’s own criteria for thinking that an infant’s representation of Spelke-objects is conceptual, we must examine how the purported object concept fares on my own criteria. Here it seems clear that the so-called object concept is not a concept because having a concept requires the ability to make judgments, and nothing the infant does seems to require an explanation in terms of judgments. What might these judgments be? Perhaps X IS AN OBJECT or X IS SOLID or X TRACES A CONTINOUS SPATIO-TEMPORAL PATH or TWO OBJECTS THERE or something like that. But if we can account for the infants’ behaviors in these paradigms without having to posit judgments like these, then why need we posit them? What use would such judgments have above and beyond various nonconceptual mechanisms, which could account for the individuation behaviors? What would be the point of making judgments, THAT’S AN OBJECT, THAT’S AN OBJECT, THAT’S AN OBJECT each time I track an object? It seems there must be some further reason which accounts for the usefulness of judgments. For example, by making assertions we can inform someone of something that we ourselves were witness to but they weren’t. In this way, we can influence another’s behavior. For example, if we are
attempting to build a camp fire, and on my wandering about I have come across some dry wood, I can assert “there is dry wood over there” and this judgment will have a clear use. Of course, I could also assert that the earth is round over and over again, but what use would this have for building the camp fire? But is it any less of a (true) judgment because it’s useless? Perhaps not, but my current point is just that (both phylogenetically and ontogenetically) there must have been some initial utility for making judgments over and above what the nonconceptual mechanisms accomplish in guiding our own behavior. It seems likely that the initial utility of judgments would have something to do with the ability to influence others’ behavior in various ways. But it seems there’s no utility in the infant judging THAT’S AN OBJECT each time it notices a Spelke-object, which rules out such judgments as good candidates for a child’s first judgments. Perhaps a better candidate would be a child’s grunting and pointing and yelling “that!” at a bottle or toy. At some point children will learn to make judgments such as THAT’S AN OBJECT, but this will be a much more abstract (and more complex) judgment that would be made only much later (perhaps in their first philosophy or physics class).

That infants are not exercising a concept SPELKE-OBJECT is clear, if one applies some of the criteria of something being a concept. The first criterion is that a concept is used, most paradigmatically, in a judgment. But, as I have suggested above, it seems there are no plausible candidates for what the purported judgments might be. In particular, if nonconceptual mechanisms (including, perhaps, visual imagery) can account

52 I suspect that imitation is also importantly involved in children learning to make judgments. I remember that when my little brother was first learning language, one of the things he did was imitate the tone and inflection of my mother’s voice, for example, when she had scolded him for getting into the cookie jar. He would point to the cookie jar and with a menacing tone proclaim “No! Not!”
for the infants’ object tracking behavior, then there doesn’t seem to be any further function that judgments such as THAT’S AN OBJECT could have. Various other criteria of conceptual thought derive from the characteristics of judgment, such as the possibility of error, and the ability to exercise concepts in isolation from the stimulus. In examining some of these it will become clear that the abilities infants display do not fit the characteristics of conceptual thought.

One of the criteria of conceptual thought which follows from the ability to make judgments is the possibility of errant judgments. The reason for this, it seems, is that judgments purport to carve up the world in certain ways (whether or not at the joints), and it seems likely that sometimes we will be mistaken about the world. Some cases of mistaken judgment are the result of misperception, such as if I were to judge that certain dog was a cat. Other cases stem from mere visual illusions, such as the judgment that the stick is bent while looking at a stick in the water. Other cases are more complex: the judgment that some people are witches, that the earth is flat, that the sun goes around the earth, etc. The fact that we can learn new things and correct our past judgments is intimately related to the possibility of error (it’s a necessary condition). We learn that the concept BENT doesn’t apply to the stick and that the concept CAT does not apply to that animal. Thus, if something is a judgment then there must be at least the possibility of making an error and, if an error is made, of correcting the errant judgment. The point about correction is important because it shows that the individual must be able to have an appreciation of his error.

Can we say of the infants in these paradigms that they have the possibility of making (and correcting) an errant judgment? From an adult’s perspective, the child is
making an error in not expecting there to be two objects when the screen is raised, such as in experiment 2 above. But, of course, the child is not making an error since the results of experiment 2 are exactly what we would expect if the child had the concept SPELKE-OBJECT. Suppose, then, (contrary to what I am arguing) that the child does have a concept SPELKE-OBJECT and that they are making judgments using that concept. From this perspective, the child makes no error in experiment 2 since the spatiotemporal information is consistent with there being only one object. Then the question is: What would count as evidence that they are misapplying the concept? It’s hard to see how one could get any evidence that the infant is misapplying the concept SPELKE-OBJECT, especially since it has been suggested (Carey and Xu, 2001, see above) that the extension of that concept be defined by the infants’ behaviors in various paradigms. For example, the fact that infants don’t track piles of sand being poured behind a screen doesn’t mean that they’ve mistakenly applied their concept SPELKE-OBJECT. Rather, it means that poured sand is not in the extension SPELKE-OBJECT. What would need to be the case for there to be evidence that the child mistakenly applied SPELKE-OBJECT? Perhaps something like that they were not surprised by an unexpected outcome when they should have been surprised (e.g. when clear spatiotemporal information was given that there are two objects, as in experiment 1). But, of course, there are some infants who fail experiment 1, since the evidence is statistical. But in that case why explain the failure in terms of a misapplication of the concept as opposed to saying the child doesn’t have the concept? So one problem is
epistemological: being able to say what the difference is between making an error with the concept and simply not having the concept at all.\textsuperscript{53}

If we assume that the child does have the concept, another problem is figuring out how the child could make a mistake applying the concept SPELKE-OBJECT. Recall, again, the graham crackers experiment where the infants chose (and crawled to) the box containing more graham crackers. If we suppose the child is making some kind of judgment like MORE SPELKE-OBJECTS IN THERE about the box with more crackers, then the question is: how might the child be mistaken in this judgment? Suppose the experimenter surreptitiously switched the number of crackers so that the box that the infant thought had more crackers actually had fewer crackers. Then suppose that once the child was shown the contents of the two boxes, it crawled to the one containing more crackers (not the one it initially crawled to). Would this be a correction of an initial judgment? It would be only if the subsequent behavior were clear evidence of a recognition that the initial judgment was mistaken. But why should we think of it as a recognition of an errant judgment rather than as a new judgment unrelated to the previous one, i.e. MORE SPELKE-OBJECTS THERE instead of (something like) NOT MORE SPELKE-OBJECTS HERE? Given an infant’s limited memory and attention capacities, it seems much more likely that the new judgment would be unrelated to the previous one, as when one distracts an infant with some new stimulus.

\textsuperscript{53} One interesting question would be how infants would fare in within-subject tests. Would success be more or less all or nothing, or would they sometimes succeed and sometimes fail the very same tests? The former would give credence to the claim that some of the infants did lack the ability, whereas the latter would raise further questions about what accounted for the differential response.
If it is characteristic of judgments that one can come to appreciate the fact that one has judged falsely, then if we need not attribute the ability to appreciate a false judgment in order to explain the data, then this should make us question whether infants are making judgments.

Another characteristic of concepts is that we can exercise them in a range of different kinds of judgments, including judgments about things that are not sensibly present. I can judge that the house is made of wood, that there is dry wood in the shed, that some people wear shoes made of wood, etc, and in each case I use the concept WOOD. Moreover, when I judge, say, that there is dry wood in the shed, I needn’t be standing in front of the shed looking at the wood; I could be driving on the freeway or sitting in the kitchen or sailing on Lake Erie. Indeed, that I can make judgments about things which aren’t sensibly present seems to be a key utility of judgments. But can infants exercise SPELKE-OBJECT in judgments about things which aren’t sensibly present? One might think that the answer to this question is yes when one considers the graham crackers experiment where the child crawled to the box containing more crackers without, however, being able to see the crackers in the box. This might suggest that children can use the concept SPELKE-OBJECT in judgments about things which aren’t sensibly present (namely, the Spelke-objects in the box). The same point would apply to the experiment in which infants reached through a slit to remove objects that had been placed within a box and were not visible to the infants. And of course there is a sense in which, as these experiments show, infants can succeed in individuating objects which they can’t currently see. And this shows that they can remember how many objects there are in the box.
I suggested, however, that this ability could be understood as a subitizing ability—an ability to compare magnitudes without knowing the exact number of things present. But subitizing is an ability that depends on perceiving something which is present, so if and infant’s ability is more like subitizing, then it would certainly lack the characteristic of concepts that they can be exercised in judgments about things not sensibly present. Of course, the infant is not perceiving the two different magnitudes when making the comparison, but they have seen the magnitudes very recently (i.e. the experimenter putting the crackers into the box). My idea is that the infant can remember these magnitudes as images of the experimenter putting crackers into the box. That the infant’s performance breaks down when comparing larger numbers, just shows the limitations of their short term memory. Analogously, suppose that I was allowed to watch men throwing wood into two different sheds at different rates, one rate being much quicker than the other. I could know which shed contained more wood without having to have made a judgment like THAT SHED CONTAINS MORE WOOD. I could know it just by comparing in memory the two images of the men throwing wood into the sheds.

If we can understand the infant’s behavior in this paradigm as a kind of subitizing ability, then we need not say they are making judgments. But suppose that the infant is making a judgment like MORE SPELKE OBJECTS IN THERE. The problem is that it is not clear what it would mean for the infant to exercise the concept SPELKE-OBJECT in certain situations where Spelke-objects are not sensibly present. For example, suppose a child were to judge SPELKE-OBJECT IN THERE, of a favorite toy in a toy box she saw two days ago. The question is: how are we to understand the content of this judgment? If a judgment is being made, it seems we should be able to say what the
content of the judgment is. But suppose there are multiple toys in the toy box. To which object does SPELKE-OBJECT refer in this case? The content of the judgment would seem to be indeterminate since we cannot tell to which object SPELKE OBJECT refers. Of course, we can make the judgment determinate (on behalf of the infant) since we can more precisely specify the content, but the question is how the child’s judgment could be made determinate using only the concept SPELKE-OBJECT.

Likewise, imagine the child judges, just before the screen is dropped but before the second object has been surreptitiously removed (as in experiment 2), SPELKE-OBJECT BEHIND THERE. To which object does this refer, the duck or the ball? In the case where a duck appears from behind the screen, we know what object SPELKE-OBJECT would refer to in the content of a judgment like THERE IS A SPELKE-OBJECT, namely, the duck that just emerged from behind the screen. But presumably the infant is also making a judgment when the duck moves back behind the screen, since (on the account being considered) this judgment is supposed to account for the infants expectations. But then what is the content of the thought THERE IS A SPELKE-OBJECT once the duck has moved back behind the screen? Is the judgment now about the duck or the ball? It seems there is no answer to this question since it could apply to either one. But if so, there is no cross-temporal integrity of the judgment: although we can say what the content of THERE IS A SPELKE-OBJECT is when either the duck or the ball is present, we can’t say what the content of THERE IS A SPELKE-OBJECT is when both objects are hidden behind the occluder.

Wiggins 1997 makes a related criticism of Xu’s idea that SPELKE-OBJECT could function as a sortal concept. His point is that SPELKE-OBJECT (glossed as “bounded, coherent, three-dimensional object which moves as a whole”) doesn’t supply sufficient
Once we move to situations even further removed from that in which the objects are sensibly present, such as judging something about a Spelke-object that the child saw at another place at another time, further problems arise. Judging things about different times and places requires one to have concepts other than indexicals, which one might think are somehow more basic than other concepts. (Even in the judgment I was imagining the infants were making in the graham crackers experiment, MORE SPELKE-OBJECTS THERE, they are still able to use the indexical THERE—of course, the concept MORE creates its own problems.) The infant couldn’t very well make a judgment about Spelke-objects in a different place at a different time without having concepts to refer to the things which aren’t sensibly present.

**How does language affect individuation?**

As noted at the beginning of this chapter, Xu thinks that learning words facilitates the formation of kind concepts. Xu wants to explain how infants are able to individuate objects by kinds, given that at first they cannot do so. Recall, for example, the ten month-old infants in experiment 1 who are unable to individuate the duck and the ball. How is it that twelve month olds are able to succeed at the same task on which the ten month olds fail? Xu believes that twelve month olds are able to succeed where ten month olds fail because they have developed kind representations, so Xu’s question is: how do identity, difference and persistence conditions such that we could answer, for any given situation, that an object at t1 is the same as the object at t2. For example, Xu’s gloss of the child’s object concept doesn’t allow us to trace the identity of objects when there is a question of whether the objects are bouncing off each other or passing through each other.
they develop these kind representations? Xu suspects that it has something to do with language and it is to this hypothesis that the following experiments are addressed.

Xu 2002 created a simple variation of experiment 2 (recounted above). The variation was that each time the object was brought out from behind the occluder, the experimenter labeled the object. For example, when a duck appeared from behind the occluder, the experimenter said, “Look, a duck!” Two different conditions where compared, a one word-condition and a two-word condition. In the one word condition, nine month old infants heard the same word each time, “Look, a toy!” In the two-word condition, infants heard a different word for each object, e.g. “Look, a duck!” for the duck and “Look, a ball!” for the ball. Results showed that infants looked longer at the unexpected outcome in the two-word condition, but not in the one-word condition, suggesting that two distinct words helped create the expectation of two objects. The results were replicated with various other nonsense words (e.g. “blicket” and “tupa”), however, when emotional expressions were used (e.g. “Ah!” and “Eww!”), infants failed to look any longer at the unexpected outcome of only one object. Xu takes this to mean that there’s something special about count nouns in facilitating individuation, as opposed to emotional expressions (Xu, 2005; Xu, in press).

---

55 Earlier in the chapter I raised serious doubt that what accounts for the success of the twelve month old infants was that they had kind representations. Nevertheless, the results of the following experiments (that words facilitate individuation) need to be accounted for within any theory of cognitive development.

56 Moreover, Xu thinks that the effect is due to the count noun itself, regardless of the fact that it was presented within a sentence (“Look, a _____!”) and the emotional expressions wasn’t (Xu, Cote and Baker, 2005, p. 376). However, she doesn’t give a reason for thinking so. I would think that if you had used an expression like “Yuk!” within a sentence like “Look, a yuk!” it would be unsurprising to get the same success as when count nouns were used.
However, Xu (in press) recognizes that the results of these experiments are consistent both with the possibility that infants are using the words as mnemonic devices and with the possibility that the infants took the words to refer to two distinct tokens rather than two distinct types. Each possibility would militate against the view that count nouns set up expectations about object kinds. Since Xu is concerned to defend the claim that count nouns are part of the causal story of how sortal concepts are learned, she needs to rule out these possibilities. The following experiments can be seen as addressing that concern.

Xu, Cote and Baker 2005 used a manual search design to test whether the use of count nouns alone (without any objects being seen) could set up the expectation of two distinct objects. Twelve month-old infants were shown a box with a slit on the front of it but whose contents the infant couldn’t see. During the two-word trials, an experimenter then looked into the box and said, “Look, a fep!” and then “Look, a zav!” three times each, making sure that the infant was paying attention after each time. Then the box was given to the infant and the infant was allowed to retrieve an object and play with it for five seconds. The object was then taken away and then the box was again pushed forward and the search behavior was monitored. The dependent variable was how long the infant would search for another object in the box. The one-word trials were identical except that only one word was used by the experimenter (e.g. “Look, a car! I see a car! A car! A car! I see a car! Look, a car!”). The rationale behind the design was that infants should search longer for another object in the two-word trials if distinct words set up expectations of distinct objects. Results showed that infants in the two-word condition did look longer than those in the one word condition, and, as with before, there was no
significant difference between the groups when emotional expressions were used. Xu suggests that these results show that it couldn’t be the case that infants are using the words as a mnemonic, since there were no objects present during the labeling phase to which the words might provide a mnemonic (in press). Rather, it seems that the count nouns by themselves set up expectations regarding the number of objects present.

But this evidence is still consistent with the idea that infants take the words to denote different tokens rather than different types. Xu, however, thinks that words are a bootstrap into acquiring kind concepts; so Dewar and Xu (unpublished, under review) carried out an experiment which attempted to show that infants do in fact take words to refer to kinds of objects. In this experiment, nine month old infants were familiarized to a box being opened to reveal either two identical items or two items which differed in color, shape and texture. In the test trials, the experimenter looked into the top of the box and described the contents with two labels (“I see a wug!” and “I see a dak!”) or one label twice. The rationale behind the design was that infants (like adults) would expect there to be two different-looking objects in the box if two different words were used, but would expect only one kind of object (even if there were two identical objects) if the same word was used twice. The results showed that infants did indeed look longer at the test trials in which two words were used and two identical objects were revealed. Conversely, when only one label was used, they looked longer at the outcome in which two different objects were shown. In addition, Dewar and Xu found the same results when only differences in shape (which is thought to be indicative of kind distinctions) were altered, but found no difference in looking time between two identical objects and two objects differing only in color (which is not thought to be indicative of kind differences). Xu takes these results as
some of the first evidence that infants do indeed expect words to refer to types of object rather than just different object tokens (in press).

This is an interesting finding, but I don’t think it shows that infants expect words to refer to kinds. At most, it shows that infants expect that different words will apply to objects that don’t look alike or are dissimilar in ways that are salient to the infant. In particular, Xu can’t claim that the fact that infants are surprised in the unexpected outcomes shows that they expected the words to refer to different kinds as opposed to merely things whose differences stand out more strongly. As noted earlier in the chapter, infants seem to find different shapes and sizes as salient differences for object individuation but not different colors. One interesting experiment to run, in light of the fact that others (e.g. Wilcox, 1999) have found that shape and size are each salient differences for infants, would be to test whether infants look longer at two different object of the same shape, but very different size. If infants did look longer at identical objects which differed significantly in size (when, say, only one word label was used—or failed to look longer when two distinct labels were used), then this would cast doubt on Xu’s claim that infants take words to refer to different kinds. Instead, it might support a different idea: that infants expect two different words to apply to dissimilar objects (not objects of a different kind). Dissimilar objects, for the infant, would just be objects that differed enough on various qualitative dimensions (or, perhaps if the difference were big enough, differed a lot on just one dimension). One would not need to refer to preexisting kind concepts to explicate “dissimilar objects.” So, for example, an infant might expect a bicycle and a bird to have different names, but not a bicycle and a car. And, perhaps,
would expect a replica toy car and the actual car to have different names (because they differ so much in size).

Xu’s general picture is that infants (around the age of twelve months) begin acquiring sortal concepts (DOG, CUP, DUCK) and that count nouns (“a dog,” “a cup,” “a duck”) help the infant bootstrap into kind concepts because infants assume that count nouns refer to kinds, and thus look for properties that would distinguish objects called by one count noun (“dog”) from objects called by a different count noun (“cat”). This picture thus depends on an underlying mechanism, independent of language, which can assemble these featural differences into representations of kinds. Words, on this view, guide conceptual development because they act as “essence placeholders,” but the “essence” (i.e. the concept) is not determined by the words themselves. Rather, there must be independent mental mechanisms that form the essence of the kind by collecting what properties things called by the same name share. Thus, Xu’s account must face the problem of abstraction. Xu doesn’t herself attempt to give an account of the mental mechanisms of concept learning, but in the last chapter (chapter 2) I reviewed the attempts of some other cognitive psychologists to provide such an account. However, I argued that they were unsuccessful in providing an account of a mental mechanism whereby concepts are learned.
CHAPTER 4: UNDERSTANDING BELIEFS

One of the most vibrant and contentious areas of research in developmental psychology is the false belief task. The basic idea of the false belief task is to determine whether a child is able to attribute a false belief to someone in predicting that person’s behavior. For example, if I know that Maxi believes that the cookies are in the cookie jar then I will predict that Maxi will look for them there rather than in the cupboard, even if I know that they are in fact in the cupboard. Predicting Maxi’s behavior in this instance involves being able to attribute to Maxi the false belief that the cookies are in the cookie jar. The standard false belief task (which was originated by Wimmer and Perner, 1983) uses just such a situation to test whether children are able to attribute a false belief in order to predict behavior.

The interesting result of false beliefs tasks is that the majority of three year olds fail the task whereas the majority of five year olds pass it (i.e. correctly say where Maxi will look for the cookie). Not only do children under five have a difficult time attributing false beliefs to others, they also have a difficult time remembering their own past false beliefs. In another version of the false belief task (the deceptive container—Perner, Leekam and Wimmer, 1987) children are shown a familiar container such as a candy box and asked what is inside. After replying that there is candy in the box, the child is shown the actual contents of the box, which to their surprise is something unexpected (e.g. pencils). Then they are asked what they originally thought had been in the box. Again, most three year olds fail this task while virtually all five year olds pass it.
There has been much debate about what the false belief task shows. Some have seen it as a kind of litmus test for a representational understanding of mind, claiming that there is a conceptual deficit between those who pass it and those who do not (e.g. Gopnik 1993; Perner, Leekam and Wimmer, 1987; Wellman, Cross and Watson, 2001). Others have tended to see the failure as due to memory and processing limitations or artifacts of the task rather than a conceptual deficit (Roth and Leslie, 1998; Fodor, 1992; Chandler, Fritz and Hala, 1989). More recently, some have suggested that the false belief task should not be seen as a crucial experiment for theory of mind since there are many other tasks which involve abilities of mental state attribution and on which children under three can succeed (Bloom and German, 2000).

Also in recent years, there has been an increasing interest in the relationship between language and success on the false belief task. The false belief task (unlike some other theory of mind tasks) involves the use of language, so it would not be at all surprising if linguistic competence were correlated with success on the false belief task. In fact, this is the case: linguistic competence is correlated with success on the false belief task, as many have reported (Bartsch and Wellman, 1995; Astington and Jenkins, 1999; de Villiers and de Villiers, 2000; Watson, Painter and Bornstein, 2002). In addition, a number of training studies have revealed that various kinds of training with language increase the likelihood of passing the false belief task (Lohmann and Tomasello, 2003; de Villiers and Pyers, 2002). Debates continue, however, amongst those theorists who view language as having an important role to play in passing the false belief task (and in theory of mind, in general). It is within this debate (i.e. among those who view language as
having an important role to play) that I will focus my attention in this chapter: How, in particular, is linguistic ability related to success on the false belief task?

**The vicissitudes of Jill de Villiers’ linguistic complements account**

In recent years, Jill de Villiers (de Villiers and de Villiers, 2000; de Villiers and Pyers, 2002; de Villiers and de Villiers, 2003; de Villiers, 2005) has defended the bold theory that children’s success on the false belief task is due to their mastering a particular syntactic structure: a complement. Complements are linguistic structures consisting of embedded propositions which (in English) follow “that” clauses, e.g., Rebekah believes *that the earth is flat*. An important feature of complements is that the complement phrase can be false and yet the whole sentence true, and it is this feature that is relevant to the false belief task since understanding false belief requires being able to grasp that it is sometimes correct to describe people as falsely believing something. Mastering a complement phrase requires that children can correctly and reliably answer the following kind of question (de Villiers 2005, p. 189): The Mom said she bought apples, but look, she really bought oranges. What did the Mom say she bought? Three year olds typically answer “oranges” while four year olds say “apples.” The empirical evidence for the theory that mastering complements accounts for success on the false belief task comes from longitudinal studies in which the best predictor of passing false belief tasks was whether there was a prior improvement with understanding complement phrases, as measured by the child’s spontaneous speech (de Villiers and Pyers, 1997; de Villiers and Pyers, 2002; Hale and Tager-Flusberg 2003). Critical to the hypothesis is the fact that none of the other measures of language acquisition was as good a predictor of success on
the false belief task as was spontaneous use of complement phrases. For example, the production of complex syntactic phrases other than complement phrases (e.g. relative clauses) did not predict success on the false belief task, nor did a measure of the mean length of utterances.

Josef Perner and colleagues (Perner, Sprung, Zauner and Haider, 2003) have raised a serious objection to de Villiers’ hypothesis about complements and success on the false belief task. He notes that if de Villiers’ theory is true—if complements are really what accounts for success on false belief tasks—then it would imply that someone who had mastered the syntax of complements would be able to pass the false belief task. But there are cases in which although children have mastered complement structure, they still can’t pass the false belief task. Perner argues that German children have complements but are unable to pass false belief tasks. There is an important difference between German and English with respect to the syntax of desire verbs: whereas in English desires are expressed without “that” clauses (e.g. Mother wants me to come inside), German typically does use “that” clauses to express desires (e.g. Mother wants that I come inside). De Villiers and Pyers 2002 claimed that complements are what distinguish beliefs from desires and that this is why desires are understood (i.e. successfully attributed) before beliefs. But as Perner points out, although German children have complements, they have just as much a problem with false belief tasks as English speaking children. Since the very same developmental gap between desire and belief (i.e., children understand desire before they understand belief) remains in German speaking children despite the fact that they use complements very early to express
Desires, it cannot be the acquisition of complements that accounts for the (later developing) understanding of belief.

De Villiers’ (2005) response to Perner’s criticism rightly backs off the claim that the belief-desire gap is accounted for by the acquisition of the particular syntactic structure of a complement, and argues instead that there is a functional class of expressions that allow for false complements and that this class must have some sort of syntactic marker. De Villiers proposes a distinction between realis and irrealis verbs. The basic idea is that the objects of realis verbs are actual objects or state of affairs whereas the objects or irrealis verbs are future or potential. For example, realis verbs such as “kick” and “carry” have actual objects or states of affairs whereas irrealis verbs like “want” and “think should” have objects or states of affairs that do not exist yet. De Villiers thinks that desire verbs like “want” are always irrealis and that the child comes to understand this from its own experiences of desiring something (2005, p. 200). But since it is the essence of desire verbs that they are about things that are not the case, desire verbs give the child no help in understanding false complements (which is the key, according to de Villiers, for being able to pass the false belief task). Rather, verbs which are able to take false complements would also have to be able to correctly describe a situation. De Villiers claims that “say” is the developmental precursor for “think” because it is realis (about something actual—i.e. the overt speech) and can therefore take

57 Do children really abstract some understanding of “irrealis” from their experience of desiring? Derek Montgomery has argued in a Wittgensteinian vein that the meanings of mental verbs cannot be abstractions from introspective experience. Instead, the meanings of words derive from their use in interactions. For example, Montgomery thinks that the fact that children never use “want” intransitively is evidence that “want” is not used merely as a label for the introspective experience of wanting. Rather, it is used transitively since it is prototypically used to get something (Montgomery 2005, pp. 115-118).
false complements since sometimes people say things that are false (ibid., p. 201). The verb “say” acts as a kind of bootstrap into the verb “think” whose occurrences, unlike “say”, are not overt. Since it is not uncommon to describe what someone said in terms of what someone thought, de Villiers proposes that the child could pick up on this (ibid., p. 202).

De Villiers 2005 proposes a developmental progression which accounts for the fact that desires are understood much earlier than beliefs. The idea is this: children’s first verbs are realis (e.g. kick, carry) followed very soon by irrealis verbs like “want” and “pretend”, whose objects are always merely potential or future. Normative verbs like “say should” and “think should” are later developing kinds of irrealis verbs. The child then comes to understand the realis verb “say” and with this understanding comes the inchoate ability to understand false complements. Finally, the child uses “say” to bootstrap into an understanding of “think.” De Villiers offers some experimental evidence to support the conclusion that children do in fact make a distinction between realis and irrealis verbs. The theory implies that children would never treat realis verbs such as “say” or “think” as irrealis. Thus a three year old should be able to correctly answer what mother thinks a child (Bella) should do but not what mother thinks Bella is doing. (“Thinks should” is irrealis and equivalent in meaning to a desire verb.) De Villiers 2005 reports preliminary data that show just this: three year old children correctly respond to questions which use irrealis verb phrases such as “want to,” “want that,” and “think should” but have trouble with questions which use the realis phrase “think that” (p. 207). De Villiers thinks that if children were not really making a realis/irrealis distinction then these results wouldn’t make sense. For example, if children treated all the questions
as realis, then they would answer “no” to any question that mentioned an event not currently true. But they don’t—they answer “yes” when asked, e.g., “Does Mom want Bella to play on the computer” although Bella is really painting a picture. Conversely, if children treated all questions as irrealis, then they would answer “no” to any question which mentioned an event which is currently true. But they don’t—they answer “yes” when asked “Does Mom think Bella is painting a picture?” and Bella is painting a picture (although mother has previously expressed that she is happy that Bella is playing on the computer).

But, supposing that children do have an implicit understanding of the realis/irrealis distinction, in what way are realis verbs that can take false complements (i.e., “think,” “say”) marked in language? De Villiers proposes that the complements following verbs such as “think” and “say” have what she calls a “point of view” markers which change the point of view from that of the speaker to that of the subject of which the speaker is speaking. For example, in the sentence “Peter thought that a unicorn was dancing in the garden,” the complement following in the “that” clause changes point of view from that of the speaker to that of Peter. De Villiers thinks that this change of point of view is unique to realis verbs like “think” and that this characteristic explains the fact that belief verbs (say, think) can take false complements whereas desire verbs cannot: “we can consider subject PoV [point of view] to be the distinctive feature dictated by this subclass of say/think realis verbs” (2005, p. 211). De Villiers also considers other syntactic features that change the point of view. For example, supposing that what Peter saw in the garden wasn’t really a unicorn but a mule with an ice cream cone glued to its head, it wouldn’t be accurate to say “Peter thought that a mule was dancing in the
garden” since this illicitly smuggles the speakers’ point of view into Peter’s. However, it does seem acceptable to say “Peter thought that the mule was dancing in the garden.” In light of this, de Villiers proposes that definite nouns within the complement phrase inherit the speakers point of view whereas indefinite nouns remain within the subjects (e.g. Peter’s) point of view.

De Villiers’ idea, then, is that point of view must be marked syntactically somehow in a language (although the syntax need not be on the surface). Thus, although English may in fact mark point of view with something like a “that” clause, German may use some other kind of syntactic marker. This allows the heart of the original proposal to be saved (i.e., the ability to solve false belief tasks requires the ability to represent false complements, and these, in turn, must be related to a unique syntactic class) while backing off the more specific claim that false complements are tied to “that” clauses. The new proposal is just that there must be a syntactic category tied to realis belief verbs that has the function of marking a point of view, and that this category of verbs uniquely allows for false complements.

The problem with this proposal is that point of view does not, in fact, seem to be unique to belief verbs, it seems to apply just as well to desire verbs. But if so, then point of view is not relevant to the realis/irrealis distinction (on which de Villiers’ account of false complements rests) and, thus, does not mark out the unique class of verbs that de Villiers would like. The problem is that de Villiers supposes that Frege problems arise only in the context of belief. But they don’t: Oedipus wanted to marry Jocasta but he didn’t want to marry his mother. But in that case, point of view applies just as much to an irrealis desire verb as it does to a realis belief verb. De Villiers invokes the point of
view marker to explain why we cannot substitute co-referring terms in belief sentences and that this property is distinctive of realis mental verbs. But the Oedipus example (which de Villiers actually uses herself, but with respect to Oedipus’ beliefs) shows that point of view also applies to irrealis mental verbs. Point of view does not map onto the category of false complements.

There is another problem with de Villiers’ account. De Villiers would have us believe that the asymmetry between belief and desire (i.e., the reason that an understanding of belief takes longer to develop than an understanding of desire) is explained by the fact that children must come to understand point of view and that point of view is marked by a unique syntactic category (i.e. realis mental verbs). However, Josef Perner and colleagues have argued that point of view is not unique to realis verbs because children have just as hard a time with attributing conflicting desires as they do attributing false beliefs (Perner, Zauner, and Sprung, 2005). But since attributing a desire in conflict with one’s own desires requires an understanding of point of view, there’s nothing special about the realis/irrealis distinction with respect to explaining children’s difficulty with false belief and the developmental gap between belief and desire. The gist of Perner’s argument is that since children understand conflicting desires no earlier than they understand false beliefs, point of view transcends the particular developmental progression which de Villiers proposes (in which point of view is characteristically tied to later developing realis mental verbs). Thus, Perner’s objection does not show that there is not a point of view, rather, it purports to show that de Villiers’ realis/irrealis distinction has nothing to do with understanding point of view.
Perner, Zauner and Sprung 2005 present a series of existing studies which, they argue, support the claim that children have difficulty attributing conflicting desires. Consider a competitive game in which you want the next card turned over to be blue while the opponent wants the next card turned over to be red. Three and four year old children in this situation typically answer incorrectly when asked which color the competitor wanted to turn up (Moore et al., 1995). Perner and colleagues argue that the difficulty for the child is that in order to attribute a correct desire to the competitor, the child must be able to recognize that the players want conflicting things and that this cannot be understood without bringing in points of view (2005, p. 235). That the difficulty stems from a point of view (and not merely a preoccupation with one’s own desires) is further supported by a study in which (in effect) each competitor has her own deck to draw from instead of only one deck. In such a condition, children’s attributions of desires to the competitor drastically improved, which suggests that the problem is one of representing differing points of view since when the differing desires can be reconciled with states of the world (e.g., I turn up a blue card from my deck and the competitor turns up a red card from his deck), the difficulty in correctly answering questions about what the competitor wants is mitigated.

It seems, then, that children’s difficulty with attributing conflicting desires results from not understanding point of view, in which case de Villiers’ revised theory faces a serious counterexample. This, in conjunction with the pretty straightforward fact that desire verbs do have points of view—as evidenced by the fact that Frege puzzles arise for desires just as they do for beliefs—makes de Villiers’ account untenable. In particular,

---

58 This study (Daxeder and Feichtinger, 2003, cited in Perner) actually used a colored die; in one condition only one die was used, in another each competitor had her own die.
these two criticisms show that her theory would not account for the fact that desire verbs are understood before belief verbs. This is so since, in effect, where de Villiers needs there to be asymmetries between belief and desire verbs, there turn out to be symmetries. Although there may be such a thing as “point of view,” it doesn’t seem to have anything to do with the syntactic categories de Villiers has proposed. Perhaps we should consider a different kind of account of the relationship between language acquisition and the false belief task.

An alternative account: The conversational account

It is possible to give an account that makes language comprehension key for passing the false belief task without thereby tying it to a particular linguistic structure. This would be preferable given the conceptual and empirical shortcomings of de Villiers’ account as well as training studies like Lohmann and Tomasello 2003 in which the most improvement on the false belief task was predicted not by training with false complements alone but in conjunction with perspective shifting discourse about deceptive objects (e.g. a sponge that looks like rock). I propose a more abstract theoretical account of the developmental lag between DESIRE and BELIEF, an account which is irreducibly linguistic but which is not wedded to any particular account of the mechanisms of conceptual-cum-linguistic development.

The idea I propose is that acquiring the concept BELIEF is a matter of learning the different uses or language games of the word “belief.” According to this account, the desire-belief gap is to be explained by the fact that the attribution of beliefs is a more complex kind of activity than the attribution of desires. Following Gauker 1994, I
suppose that the most basic type of linguistic act is a command. A child is undoubtedly
familiar with commands due to the fact that many of the interactions between the child
and the caretaker will involve commands such as “don’t go in the street!” and “no
cookies before dinner!” Children learn how to obey commands and would thus have an
early developing understanding of them.

To attribute a desire is, in effect, to give a command on another’s behalf (Gauker
1994, 2003). So, for example, an older sibling or the father might say “Mother said not
to eat any cookies before dinner” and we might expect that the child would learn to obey
such a command as she would were the mother herself issuing it. Commands have direct
consequences for behavior, and in recognizing this a child who learned to attribute a
desire to an acknowledged authority figure (like mother) would have at her disposal a
tool to make people do things. This is, at root, the “cash value” of desire attribution, and
it is this kind of desire attribution that we might expect to emerge very early in mental
state attribution.

In contrast, the attribution of a belief is an assertion on another’s behalf (Gauker
1994, 2003). The prototypical kind of situation in which attributions of belief would
occur would be that in which people are attempting to solve a problem and are collecting
information. For example, suppose a couple of three-year-olds, Sam and Bill, are playing
a game in which they have only one chance to guess where the cookies are (which they
are then allowed to eat). Then suppose that Sally walks in and whispers to Bill that the
cookies are in the oven, and then leaves. (Suppose also that Bill has some reason to
believe Sally—or at least no reason to disbelieve her.) Bill would then need to tell Sam
that Sally said that the cookies were in the oven, and in doing so he would attribute a belief to her. This is the kind of situation in which belief attribution might arise.

But the attribution of a false belief would take something more: rejecting an assertion made on another’s behalf. So, sticking with our scenario, suppose further that Sam knows that the cookies are not in the oven because he can see in the oven (although Bill can’t). In such a situation, Sam would be in a position to reject Bill’s assertion on Sally’s behalf, and if he were to do so he would be attributing to Sally a false belief. This is the kind of situation in which the attribution of a false belief would arise.

Although it doesn’t seem implausible that Sam could reject Sally’s assertion, it does not seem as likely that Sam would be able to convey this to Bill with any success. Sam would have to acknowledge that Bill really was told by Sally that the cookies are in the oven (as opposed to simply rejecting what Bill said) and also that Sally really does believe that they are in the oven (as opposed to calling her a liar), but maintain that what Sally thinks is mistaken. It is difficult to image two three year olds navigating a conversation of this complexity. Thus, given that the attribution of a false belief would characteristically involve a kind of collective problem solving, and that such collective problem solving seems beyond ken of a three year old, it would follow that attributions of false beliefs are not within the purview of young children.

According to the view I am proposing, the developmental lag in understanding the concept BELIEF is due to the fact that attributions of beliefs are part of a more complex language game involving a kind of collective problem solving. The young child’s lack of understanding of the concept of belief is due to the fact that they have not yet learned how to navigate such collective problem solving situations. This makes the
acquisition of the concept BELIEF strongly dependent on language since to acquire the concept BELIEF is just to be able to successfully navigate the nuances of the language game of “belief.” Importantly, for the purposes of this chapter, this includes being able to attribute false beliefs which, I have suggested, is just to be able to reject an assertion made on another’s behalf when engaging in a collective problem solving task.

Applied to the standard false belief task—the change of location task discussed at the beginning of the chapter—the account I am proposing would say that the ability to attribute to Maxi a false belief requires the ability to assert what Maxi will say on Maxi’s behalf, even though one knows it isn’t true. The experimental setup would be viewed as a situation where there is a problem to be solved and the experimenter is attempting to pool information in order to solve the problem. The failure of younger children is explained by the fact that they do not yet understand how to navigate a collective problem solving task like this, while older children, on the other hand, have learned what is required in these situations: they have learned how to assert something on another’s behalf.

It may be useful to contrast my proposal with the two main types of account of three year olds’ difficulty with false belief in the literature: performance accounts and conceptual competence accounts. According to performance accounts (Fodor 1992; Chandler, Fritz and Hala 1989; Chandler and Hala 1994; Hala and Chandler 1996; Leslie and Roth 1993; Bloom and German 2000; Onishi and Baillargeon 2005; Birch and Bloom 2004), the three-year-old’s difficulty is not that they lack the concept BELIEF, but that the false belief tasks overtax the three-year-old’s limited information-processing systems. Such accounts attempt to show how altering certain factors increases
performance on false belief tasks. Factors which have been shown to increase performance include things like framing the task in terms of explicit deception or trickery, involving the child in actively making the key manipulations (e.g. the child moves the location of the object), and highlighting the salience of the protagonist’s mental state and/or reducing the salience of the actual state of affairs. Inasmuch as these manipulations enable younger children to be successful at false belief tasks, the failures are thought to be due not to a lack of the concept BELIEF but, rather, to the fact that the conceptual competence is masked by the difficulty of the task. Making the tasks easier unveils the child’s underlying conceptual competence.

Conceptual competence accounts, on the other hand, attempt to show that the difficulty with false belief tasks is not reducible to performance limitations and so must be due to an altogether lacking or different concept of belief (Wellman, Cross and Watson 2001; Gopnik and Astington 1988; Gopnik 1993; Wimmer and Hartl 1991; Perner 1991; Perner, Leekam and Wimmer 1987). De Villiers’ theory is a kind of conceptual competence account since she thinks that young children’s performance on the false belief task exhibits a leap that cannot be accounted for by performance factors alone. Nevertheless, she contrasts her theory with a conceptual competence account such as Perner’s according to which the developmental change occurs independently of particular linguistic competencies (2005, p. 199).

The theory I am advocating agrees that there is a new competence that children learn and that it is not already there but merely masked by performance limitations, as the
Nevertheless, the abilities required for children to be able to successfully navigate the belief language game are affected by the kinds of information/processing competencies—a point which even the conceptual competence accounts can accept. Furthermore, I accept that a child’s understanding of the language game of “belief” is mediated by the kinds of information-processing and memory abilities that the performance account claims accounts for the eventual success on false belief tasks. The difference is that the performance account claims that all along the child has a concept of belief, whereas I am claiming that the child only acquires this concept by mastering the linguistic interchanges characteristic of collective problem solving endeavors. The theory I am advocating also disagrees with de Villiers’ separation between a conceptual matter and a linguistic one since according to my theory there is no conceptual competence outside of linguistic competence.

**Objections to the conversational account**

The conversational account of belief faces a couple of prima facie problems. First, it might be thought to imply that attributing a false belief should be no more difficult than attributing a true belief. This is so because belief attribution is more difficult because of the more complex language game than that of desire. But children’s difficulty is purportedly only with attributing *false* beliefs. Second, if prelinguistic children can be

---

59 Wellman, Cross and Watson 2001 argue that performance accounts must not only show that altering the task changes performance, they must in addition show that such alterations differentially affect younger children. This is additional burden of proof must be met since the performance account claims to explain the differential performance of younger children who fail the false belief task and the older children who pass it. Wellman and colleagues argue that the studies do not meet this burden.
shown to understand false belief then the account faces a serious counterexample. I will consider these in order.

According to my account, an understanding of belief requires one to have much more sophisticated interpersonal linguistic skills than does an understanding of desire. To understand belief requires that one be able to assert something on another’s behalf as well as to reject such an assertion; and to do this is to be able to engage in a kind of conversation or debate. To understand desire, on the other hand, requires only that one be able to carry out and issue commands on another’s behalf. One possible objection to this idea is that it should imply not only that children should understand false beliefs later than desires but also that they should understand true beliefs later than desires. If belief attribution requires more sophisticated conversational skills than desire attribution then we should expect any belief attribution to lag behind desire attribution. In short, young children should have difficulty attributing true beliefs without a corresponding difficulty in attributing desires.

One of the problems in being able to say that children can attribute a true belief is distinguishing between the tactic of merely repeating what is in fact true (when answering a question about a protagonist’s belief) and the tactic of attributing a belief. If the cookies are really in the cookie jar and we ask a three year old where Maxi thinks the cookies are, the answer “in the cookie jar” is compatible with either tactic. So we can’t say that three year olds understand true beliefs (but not false ones) based on such evidence. This consideration led Riggs and Simpson 2005 to pose questions about past true beliefs to three and four year olds. Riggs and Simpson used an unexpected transfer task (Wimmer and Perner 1983) but added a condition in which subjects were asked
about a protagonist’s past true beliefs. For example, consider a story in which the protagonist, Linda, is reading a book in the living room and then goes to the kitchen for a drink of water, leaving the book in the living room. While in the kitchen, Linda’s sister comes in and takes the book upstairs to read. A past true belief question would be: When Linda left the living room, where did she think her book was? In contrast, a false belief question would be: Where does Linda think her book is? Riggs and Simpson found that three and four year old children fared equally poorly on both questions, suggesting that past true beliefs are no easier to attribute than false beliefs (2005, p. 29). Importantly, these same children’s memory for past states of affairs in the world (which are no longer the case when they are asked the question) remained near ceiling, showing that it isn’t merely a problem of recalling past states of affairs which are not currently true but, rather, a difficulty with attributions of belief (ibid.).

This evidence is consistent with the conversational complexity account which would predict that an understanding of belief, including true beliefs, would come later than an understanding of desire. What about currently true beliefs? Can young children attribute currently true beliefs? But what is to say that they do not use the tactic of simply answering based on what is the case? It is, of course, possible that young children could attribute beliefs that are currently true rather than using the tactic of answering the belief questions by answering what is in fact the case. One reason to doubt such a claim is the fact that children’s spontaneous speech about beliefs doesn’t emerge until after spontaneous talk about desire (Bartsch and Wellman 1995, esp. pp. 96-102). This, taken in conjunction with the evidence of young children’s other difficulties with belief, should make it seem somewhat ad hoc to say that, nevertheless, children are able to attribute
currently true beliefs in situations where the child could easily answer correctly just by saying what is the case.

A second objection to the conversational complexity account is the purported evidence that very young children (15 months) and possibly chimpanzees can attribute false beliefs in predicting behavior. If true, this would present a serious problem for the conversational complexity account since it claims that to attribute a belief just is to assert something on another’s behalf in a conversation. This implies that belief attribution does not occur outside of linguistic practices. If prelinguistic children (or nonlinguistic chimps) really are attributing beliefs—even false beliefs—then the conversational complexity account is false.

Onishi and Baillargeon 2005 used a violation of expectation paradigm to test whether fifteen-month-old infants expected someone to look for an object where they (the actor) believed it to be (rather than where it actually was). The basic idea of the experiment was to hide an object in one of two locations (in a green box and in a yellow box) and under different conditions (i.e. when the actor saw the object being hidden vs. when the object was hidden or moved while the actor was absent) and then record the amount of time the infant looked when the actor subsequently looked for the object in the different locations. The prediction was that infants would expect the actor to look for the object where the actor believed it to be. So, for example, the infant should look longer when the actor looked for the hidden object inside any box in which they did not last see it. Onishi and Baillargeon did find longer looking times for conditions in which the actor looked for the object in the “wrong” place (i.e. some place other than where they last saw it), and they take this to show that “15-month-old infants…realize that others act on the
basis of their beliefs and that these beliefs are representations that may or may not mirror reality” (2005, p. 257). If this is true and prelinguistic infants really do attribute beliefs to others in order to predict their behavior, then the conversational account of belief, according to which belief attribution is the purview of linguistic contexts, would face a devastating counterexample.

The basic line of my response to this and other purported evidence of a prelinguistic belief concept (e.g. O’Neill 1996) is to claim that while there may be mental mechanisms which are able to keep track of certain contingencies in the environment and which lead the infant to form certain expectations, these abilities do not entail that the infant is attributing beliefs. Rather, the infant’s expectations are to be accounted for in terms of behavioral associations (which may be either learned or innate) such as that people will look for an object where they last saw it. Such expectations need not involve reference to mental states but, rather, associations between visible events such as someone’s looking somewhere and someone’s searching for something (c.f. Perner and Ruffman 2005; Povinelli and Vonk 2004; Andrews 2005). If someone’s seeing something in a particular location is associated with someone’s searching for something, then when this association is broken the infant would look longer at that event.

Onishi and Baillargeon consider such a tactic but reject is because of considerations of parsimony. In particular, they think that it requires attributing to the infant overly complex machinations:

Perhaps the infants brought to the task a superficial expectation (acquired through repeated observations) that a person looking for an object will search for it where she last saw it disappear. This interpretation…assumes that the infants (i) distinguished between their own and the actors perceptions; (ii) kept track of what the actor did and did not see; and (iii) understood that the actor’s perceptions
Onishi and Baillargeon claim that two parsimony-related considerations tip the scales in favor of their belief attribution interpretation of the results over the association of perceptions and actions interpretation. First, they think it likely that an understanding of mental states is innate (perhaps modular) and so children already have the concept of belief, they just have to learn when to apply it. Second, they think that other studies that show that infants can “predict where an actor will search for a hidden toy even when she does not see it disappear but must infer its location based on various cues” makes it unlikely that infants are using “an extensive series of superficial expectations linking different perceptions to different actions” (2005, p. 257).

But is it really such a complex thing for there to be associations between looking behaviors and searching behaviors? I don’t see why this would be any more “sophisticated” than attributing beliefs; indeed, it strikes me as less sophisticated. On the one hand, infants may possess an innate association between characteristic looking behaviors and behaviors such as approaching the object, investigating the object, and searching for the object. Such associations would lead them to expect that if someone is looking at something then they will subsequently search/investigate the object. On the other hand, perhaps infants learn to associate looking behaviors and searching/approaching behaviors and consequently acquire certain related expectations. (Onishi and Baillargeon seem to assume this construal of the alternative account in their discussion.)

Regarding the conditions needed to learn such an association, (i) seems in no way problematic: someone’s looking at something has a characteristic kind of appearance
(e.g. head turned in certain way, eye gaze fixed, etc.) and an infant’s perception of this would not be confused with their own lookings. Moreover, to say that these two things would be confused seems to assume that the process of association is one in which infants consciously engage. But it is no part of the behavioral association account to say that this is a process in which infants consciously engage. Regarding (ii), it seems clear that infants can keep track of certain kinds of lookings: if the person is looking at the toy then one might expect them to approach or investigate that toy. On the other hand, if there was no one there to see the toy moved to or hidden in a different location, then since no association between perception and behavior exists, the infant would not expect any further behaviors regarding the object in that location. They might instead default to where the last association was made. Regarding (iii), another way to put this is just that infants as a matter of fact have certain expectations as the result of certain (innate or learned) associations between kinds of behaviors. Again, the infant needn’t have to consciously “figure out” any of this since the association processes are subpersonal.

Regarding Onishi and Baillargeon’s parsimony considerations, one cannot appeal to parsimony by assuming an answer to the very thing that is in question, namely, whether or not prelinguistic children attribute beliefs. Of course, if it was established that belief attribution is a modular process, then they could appeal to parsimony since it would be accepted on independent grounds and they could then invoke it as an explanation of the current data. However, whether belief attribution is a modular process (as theorists such as Scholl and Leslie 1999b and Fodor 1992 claim) is not a datum to which one can appeal in making a parsimony argument.
Perner and Ruffman 2005 offer another interpretation of Onishi and Baillargeon’s data which does not require that the infants can engage in belief attribution. The idea is that infants make an association between the actor, object and location of the object in the familiarization condition and that the test conditions in which these associations are more similar to that of the familiarization condition require less processing and therefore a shorter looking time. In particular:

During the familiarization phase of the test, an adult actor watched by the infants last observes the object (watermelon slice) in the yellow box under two conditions and in the green box under another two conditions. Neurons remember this information both in an active manner (through sustained firing in the prefrontal cortex) and in a latent manner (through altered firing thresholds in nonfrontal regions). If an association of elements “actor-object-yellow box” is still sustained in the frontal cortex when babies are exposed to the test stimuli, a consistent test combination [i.e. an association more similar to the original one presented] will need less processing and, consequently, a shorter “looking” time than a new combination of elements (e.g., actor-object-green box). In the latter case, babies might need longer looking times because, when they examine the new combination, they must form a new association (Perner and Ruffman 2005, p. 215).

This proposal is more deflationist since it doesn’t even require that infants come to the experiment with already-stored behavioral associations, but that the associations are made in the context of the experiment. This may be preferable given that older children (e.g. two and three year olds) fail other nonverbal false belief tests. Either way, the interpretation of Onishi and Baillargeon’s findings can be accounted for without supposing that the infants are engaging in anything like belief attribution.

The possibility that 15-month-old infants could be tracking certain behavioral regularities raises the possibility that even four and five year olds are succeeding in the verbal false belief tasks not because they successfully attribute beliefs but because they have learned a behavioral regularity. What, then, counts as evidence that one has the
ability to attribute beliefs? On the account I am offering, an understanding of belief requires that one can successfully navigate the intricacies of the language game of belief. Importantly, they must be able to reject an assertion made on another’s behalf. Children who can successfully navigate such linguistic exchanges count as having the concept of belief. This leaves unresolved at what point the nascent language user’s success on various false belief tasks is due to representing behavioral regularities such as “people look for things where they last left them” or actually understanding the language game of “belief.” Success on nonverbal versions of the false belief task, I claim, are due to behavioral regularities, whereas success on versions which are heavily linguistic are due to mastering the language game of belief.

Kristin Andrews has recently argued that any paradigm which purports to test for an understanding of mental states but which uses only behavioral prediction as the dependent variable could always be explained in terms of an understanding of behavioral regularities rather than an understanding of mental states (Andrews 2005). Andrews’ subject matter is research on chimpanzee theory of mind, although she takes her point to apply equally to paradigms in developmental psychology which use only behavioral prediction as the independent variable (2005, pp. 528-530). What is needed to show a genuine understanding of mental states, Andrews thinks, is not merely behavioral predictions but, rather, the search for mentalistic explanations, a paradigm case of which would be the investigation of anomalous behaviors. Suppose we want to know whether a chimpanzee can attribute the concept SEEING to others. In order to test for this, it is not sufficient just to show that a chimpanzee who has experienced wearing an opaque bucket will not beg for food from an experimenter wearing that same opaque bucket, since the
chimp need only attribute to the experimenter the inability to do things rather than the inability to see things. Andrews proposes instead the following kind of experiment: Let the chimpanzee wear the opaque, non-seeing bucket but then let it watch one of its conspecifics wear the same non-seeing bucket but behave as if it could see (because the opaque, non-seeing bucket has been surreptitiously swapped for an externally identical one of the seeing variety). The dependent variable should be whether or not the chimpanzee then goes to investigate the cause of this anomalous behavior, e.g. attempts to put on that bucket to determine whether it can see out of it.

Andrews is right, I think, to emphasize that behavioral prediction paradigms are not sensitive enough to discriminate between the behavioral association tactic and mental state attribution. And perhaps the investigation of anomalous behaviors such as the experiment she proposes would give us further reason to believe that the chimpanzee really does attribute mentalistic phenomena. However, in the case of infants we have another way of being sure the infant has a concept of belief. The fact that people talk about mental states is a much shorter and surer route to knowing that they engage in mental state attribution.

To reiterate: my response to the challenge that prelinguistic infants (or nonlinguistic creatures for that matter) attribute mental states in predicting behavior is to acknowledge that there are likely certain tactics that account for the behavioral predictions but that these tactics do not require mental state attribution. In general, I want to urge a dissociation between the prediction of behavior (in experimental paradigms such as Onishi and Baillargeon 2005, which don’t use language) and the appropriate use of mental terms like “believe,” “want,” and “think” (in the verbal versions of the false
belief task). The former does not require a mentalistic interpretation and is not dependent on language; the latter does require mentalistic interpretation (if anything does) and is dependent on language.

Other supporting data for the conversational account

The conversational account of success on the false belief task is able to accommodate a number of other findings. One important source of information regarding the development of an understanding of mental states comes from research with children with autism as well as deaf children (in particular, deaf children who do not have parents who use sign language). Both groups are significantly delayed in passing theory of mind tasks such as the false belief task and yet unimpaired in intelligence and other developmental measures—a point which provides support for a conceptual deficit (Baron-Cohen, Leslie, and Frith, 1985; Baron-Cohen 2000; de Villiers and de Villiers, 2000; P. de Villiers 2005). Furthermore, these developmental deficits have been strongly correlated with impairments with language, and further longitudinal studies have shown increased success on language ability to precede success on false belief tasks (Tager-Flusberg and Joseph 2005; P. de Villiers 2005). Theorists who believe that language plays an important role in theory of mind development have taken this evidence as strong support for their position. These same theorists, influenced by Jill de Villiers’ theory (outlined earlier), have made the stronger claim that the acquisition of complement phrases is primarily what accounts for later success on false belief tasks (de Villiers and de Villiers, 2000; Tager-Flusberg and Joseph, 2005; P. de Villiers 2005). However, in light of the criticisms of that theory presented earlier, the complement theory seems
untenable. My own proposed account would predict the developmental lag seen in both autistic children and deaf children based on their difficulties with language. In particular, since these children, for one reason or another, do not participate in conversations, their ability to attribute beliefs—which on my proposed account is, at root, the ability to assert something in conversation on another’s behalf—would be systematically impaired.

An interesting question regarding the false belief task is whether children can succeed in attributing a past false belief to themselves before they can succeed in attributing a false belief to others. There is evidence that suggests that young children are no better at one than the other and that knowledge of one’s own false beliefs doesn’t precede knowledge of others’ false beliefs (Gopnik and Astington, 1988; Gopnik 1993; Gopnik and Meltzoff, 1994; Wimmer and Hartl, 1991). The results of the meta-analysis carried out by Wellman, Cross and Watson 2001 confirms this finding. These results could be taken as support for the theory-theory of belief and as evidence against the simulation theory. The reason would supposedly be that if the concept of belief is a theoretical entity which is constructed to account for certain observable data, then it would be equally applicable to both oneself and others, in which case one wouldn’t expect it to be applied to oneself any earlier than to others. In contrast, the simulation theory would seem to require that one can understand one’s own beliefs before one can attribute beliefs to others since on this account we attribute beliefs to others by simulating them using our own beliefs. Thus, the simulation theory would seem to require that our access to our own mental states precedes our successful attribution of beliefs to others.

However, this developmental trend of a parallelism between self-knowledge and knowledge of others does not only support the theory-theory, it also supports the
conversational account of belief. In particular, there would be no particular reason on the conversational account to expect that the attribution of beliefs to self would be any easier than the attribution of beliefs to another. Asserting something on one’s own behalf which one now knows to be false would occur in just as complex a situation (i.e. a collective problem solving situation in which information is being collected) as asserting something false on another’s behalf. But if the complexity of the situation for belief attribution to oneself is just as great as for belief attribution to others, then we wouldn’t expect children to succeed any earlier on false belief tasks for self than for other.

Conclusion

Psychologists who attempt to provide a fundamental role for language in the acquisition of mental concepts are right to try to do so. Jill de Villiers’ complements theory is a brave attempt to tie successful mental state attribution to language. However, it fails because it ties language too strongly to particular linguistic structures, as Perner’s rebuttals have shown. Moreover, it is hard to see how a merely syntactic aspect of a language can itself account for new semantic properties of thought. If language is to bring about new conceptual abilities then an account must be given of how this is done. The conversational account suggests that to acquire a new concept—in particular, mental concepts such as BELIEF and DESIRE—is just to learn the uses of words like “believe” and “think” and “want.” Language creates new semantic properties—the contents of concepts like BELIEF and DESIRE—by being used by speakers in a certain way. I have not, of course, attempted to offer such a semantic account. Rather, my goal in this chapter has been to suggest that if we really want to take seriously the idea that language
plays an important role in cognitive development, then we might do well to look at a
different account of how language can play such a role. The conversational account is
the beginnings of such an account; there remains much to be developed. But I hope to
have at least shown how it can handle some key data from the false belief paradigm as
well as other developmental data.
In this chapter I look at four philosophers who attempt to provide a role for language in making possible new kinds of thought: Andy Clark, Daniel Dennett, José Bermúdez, and Peter Carruthers.

**Andy Clark: Cognitive offloading**

Andy Clark (1997, 1998, 2002) has developed a thesis about the role of language in cognition which purports to make language the fundamental difference between the cognitive abilities of nonlinguistic creatures, such as apes, and linguistic creatures, such as humans. Clark has many irons in the fire, including a defense of his version of the so-called extended mind thesis, and one cannot really understand his account of the role of language in cognition without understanding his account of representation, which is integrally connected to the notion of embodied cognition. My aim here is not to give an exhaustive review of all of Clark’s work relating to representation but, rather, to focus on the particular role that he sees language as playing in cognition.

I think the best way to understand Clark’s whole thesis is in terms of the notion of offloading problem solving work onto the environment. In order to understand what offloading might mean, consider a contrast between two different ways of solving a particular problem in robotics. Suppose we wanted to build a robot that would pick up empty pop cans that were scattered about. One way to solve this problem would be to create a robot that represented the environment via a detailed map which it updated as it
moved around. It would isolate a pop can (perhaps by its distinctive shape) and then plan a route to pick up the can (avoiding any obstacles represented by its map of the environment). Such a solution to this problem domain would be representationally and computationally rich because it would involve detailed representations (e.g. the map, the route to the pop can) and manipulations of these representations.

Contrast this “classical” solution with a solution that Rodney Brooks’ robotics lab came up with: Herbert (the following is cited in Clark 1997, p. 14). Herbert employs several simple routines that are arranged in a “subsumption architecture.” The basic idea is that you give the robot some simple routines to follow and these routines are interrupted when some specified property is detected, at which point a new routine is begun. For example, Herbert is given several basic routines such as an obstacle avoidance and locomotion. Once a table-like shape is encountered (using a simple vision system), the locomotion and obstacle avoidance routines are turned off (e.g. by negative feedback as a result of registering the table) and new routines are turned on (e.g. by positive feedback). These included: sweeping the table with a video camera until a can shape was encountered, rotating itself once a can was found so that the can was in the center of its field of vision, extending its arm until a can shape was encountered and, finally, grasping the can. Each of these routines would be initiated by positive feedback initiated by a particular event (e.g. centering the can in the field of vision, encountering a can with the sensors on its hand, etc.).

---

60 One well-known problem with updating internal representations is the frame problem. This is the problem of figuring out which environmental changes are relevant to update, since updating all of them is time-consuming and computationally expensive (McCarthy and Hayes, 1969; Dennett 1981).
The latter solution, Herbert, differs significantly from the former in that it involves no centralized control and thus no need for the different routines to be able to communicate with each other (in some kind of internal code). Rather, all that is needed are some relatively simple and self-contained behavior routines which are controlled by their interaction with the environment. No centralized controller is needed other than the task-specific mechanisms and some simple feedback links between them which are controlled by environmental stimuli.

Another example: Imagine that you need to design a robot that will navigate itself to a charging station, which is indicated by a light source, and located between two poles. One solution would be to enable the robot to compute a trajectory between the two poles. Such a solution would be representationally and computationally rich because it would involve representing each of the poles as well as its own position in relation to the poles, and then using these representations to compute a trajectory between the poles. Contrast this solution with a robot that used two behavior routines: 1) a phototaxis system which zigzags towards any light source and 2) an obstacle avoidance mechanism which turns away when it bumps into something. These two behaviors would solve the problem because the robot would be continually attracted to the light but if it ran into one of the two poles it would turn away but then turn back towards the light again, this time from another angle. Although such a solution would involve more trial and error than the former, it would be much simpler representationally and computationally, involving only two, self-contained routines connected by some sort of simple feedback. And what makes it simpler seems to be that it interacts with and exploits its environment in coming to a solution in a way that the trajectory-computing robot does not. The phototaxis robot
would not have to compute anything. All it has to do is follow the light source (an ability that even plants have); and it “navigates” obstacles simply by bumping into them—it doesn’t have to represent them beforehand or figure them into the trajectory it computes (the preceding example is cited in Clark, 1997, p. 108).

Cognitive offloading is the idea of exploiting certain features of the environment in problem solving. Both Herbert and the phototaxis robot exemplify this notion of cognitive offloading, in contrast with the solutions that posit the manipulation of representations in some sort of inner code or language. There is no end goal that is represented and which guides the behavior of the offloading robots, as there would be with, e.g. the robot that represents both the starting position and the ending position and then transduces those coordinates into the physical movements of the robot. In this case, the robot’s physical movement is indeed guided by a representation of a goal state. Contrast the representation of the light source in terms of some coordinate system with the phototaxis robot, which does not seem to be guided by an internal representation at all (at least, no more than the sunflower is). The phototaxis robot is controlled directly (i.e. not mediated by a representational economy) by the light source as the robot moves around in its environment. Likewise, Herbert moves more or less randomly around in its environment until it encounters some specified object.

The idea of a solution depending on moving around in and interacting with the environment is an important idea for Clark. Classical cognitive science, he thinks, has been misled by the kinds of problem domains with which it has been concerned, such as playing chess and generating past tense forms of English verbs (Clark, 1997, p. 67). Dealing with abstract problem domains has led to a neglect of the essential action-
oriented perspective of real-life organisms: these cannot afford to sit back and waste time computing a solution. Rather, what is needed are simple mechanisms that perform their tasks quickly by using minimal information processing (Clark, 1997, pp. 7, 46-47).

Clark’s view of the cognitive agent is action-oriented and this contrasts with more classical cognitive science views according to which agents are seen as passive compilers of information about their environment.

The notion of offloading leads Clark to an understanding of representation as much sparser. Instead of having to posit detailed inner representations interacting with each other, the organism is guided by a collection of some very simple mechanisms which are tuned to interact with their respective environmental contingencies. The light-seeking robot, for example, uses a simple mechanism which blindly moves towards a light source; and a simple obstacle-avoidance mechanism uses the presence of some environmental stimuli (e.g. bumping into something) to cause some behavioral routine in the same way that a thermostat clicks on the furnace once the temperature drops below a specified measure. These mechanisms are simple because we can see them as either “mere causal correlations” or as “adaptive hookups” (Clark, 1997, p. 147). Clark is fairly pragmatic about his representationalism, viewing various mechanisms as existing on a continuum, where something is more or less representational relative to the explanatory benefits we would gain from treating it as such (1997, pp. 147, 152-153, 168).

Representation talk gets its foothold, I suggest, when we confront inner states that, in addition, exhibit a systematic kind of coordination with a whole space of environmental contingencies. In such cases it is illuminating to think of the inner states as a kind of code that can express the various possibilities and which is effectively “read” by other inner systems that need to be informed about the features being tracked. Adaptive hookup thus phases gradually into genuine internal representation as the hookup’s complexity and systematicity increase (Clark, 1997, p. 147).
In the cases where we find positing internal representations explanatorily useful, they will be useful because we will have been able to “unpack the complex causal web of influences so as to reveal the information-processing adaptive role of some system of states or of processes” (ibid., p. 174). Clark seems to subscribe to a teleological account of representation (c.f. Millikan, 1984) according to which representational content is explicated in terms of the adaptive function that some particular inner state plays (ibid., p. 146). Representation for Clark thus turns on two issues: 1) the complexity of the environmental contingencies tracked by some internal state and 2) the functional role that the content plays in relation to other internal states (ibid.). The explanatory utility of positing representations, for Clark, derives from his belief that to truly explain something is to be able to build it (ibid., pp. 117, 120).

Thus, Clark’s real target is not representationalism itself but, rather, a certain kind of representationalism which, as a result of neglecting the interaction of the action-oriented organism with the environment, posits overly rich, internal representations. For Clark, there are two kinds of situations which are, what he calls, “representationally hungry” problems—those for which positing some sort of representation seems required: the ability to think about non-existent or absent properties as well as abstract or unruly properties (ibid., p. 167).

In the two ranges of cases (the absent and the unruly), the common feature is the need to generate an additional internal state whose information-processing adaptive role is to guide behavior despite the effective unfriendliness of the

---

61 This is why Clark is not ultimately satisfied with the kind of explanation given in Dynamic Systems theory. DST, although being able to predict counterfactuals through various equations, doesn’t give us any clue about how to reconstruct or build the system, nor does it help us to explain specific deficits in cases of localized damage (Clark, 1997, pp. 117, 126).
ambient environmental signals (either there are none, or they require significant computation to yield useful guides for action). In these representation-hungry cases, the system must, it seems, create some kind of inner item, pattern, or process whose role is to stand in for the elusive state of affairs. These, then, are the cases in which it is most natural to find system states that count as full-blooded internal representations (ibid., p. 168).

That Clark would see these kinds of cases as “representationally hungry” makes sense given his view of action-oriented, embodied cognition which I have briefly sketched here. The gist of that view is that we can use the environment to guide our behavioral responses, so when it comes to the question of how we guide our behavior in the absence of such cues, there is a prima facie gap that needs to be filled. For example, although we can explain how the phototaxis robot guides its behavior by continuous interaction with its environment, we cannot explain how, for example, my desire to avoid person x (who I believe is in an adjacent room) leads to my exiting the building through an alternate route, or why I vote Democrat. This is why Clark needs to talk about some other kind of “stand in” for the environment in cases where such environmental guidance is lacking.

Clark and Grush 1999 give an example of a “representationally hungry” problem: understanding how we can perform fluid reaching behavior, an ability that depends on “the brain’s receiving and responding to a stream of proprioceptive feedback” (p. 6). The obstacle in explaining how we accomplish this is that the feedback that would be provided by real-world, interactive, proprioceptive feedback would be too slow to be able to guide the reaching behavior. One potential solution is to use an “emulator” to model and predict the next state of the system. Such an emulator would take into account various states of the system (e.g. changes to shoulder and elbow angle, changes in the angular velocity and acceleration of the elbow joint, etc.) as well as the current control command issued and yield a prediction of the next state of the system (ibid.). Emulators
count as representations precisely because they are “inner items capable of playing their roles in the absence of the on-going perceptual inputs that ordinarily allow us to key our actions to our world” (ibid., p. 9; c.f. Dennett 1991, p. 191).

The role of language in cognition

The same theme of offloading problem solving onto the environment can be used to understand the way that Clark thinks about the impact of language on cognition. Offloading to the environment enables us to solve problems while using less complex representational and computational resources. So too, Clark argues, more complex forms of cognition arise because humans (and perhaps some hominids) are able to create cultural and institutional artifacts that vastly increase their problem solving abilities. The increased problem solving capacities of humans, Clark argues, are not due to humans having much more complex representational systems but, rather, to our ability to create and arrange our physical and social environment in such a way so as to get the most punch out of our short-sighted, relatively dumb, pattern-completing brains (1997, pp. 179-180). Essentially our minds create and maintain an external scaffolding which, when our naked minds are embedded in them, accounts for the particular brilliance of human minds (ibid.). Clark thinks that the most important of these scaffoldings for characteristically human cognition is language (1997, 1998, 2002).

A simple example of how we “scaffold” our brains is the use of external memory aides. For example, we can learn to recognize to write “49” when we see “7 x 7” and this is just the kind of task at which individual pattern completing brains excel. But although, I suppose, we could memorize larger and larger multiplication tables, there is no reason
we need to if we have a pencil and paper and have been taught how to break down a problem like “777 x 777” into simpler steps. Essentially, we transform the complex problem into simpler problems using the pencil and paper. This is a simple case of what Clark calls “external scaffolding”: the pencil and paper enable us to confront simpler, more tractable pieces of the more complex problem, and hence become an integral part of our cognitive feats. In this case, the pencil and paper function as a scaffold to our limited working memory: we can store the intermediate solutions on the paper instead of having to try to remember them—a task which, perhaps, only the most gifted mnemonicist could pull off.

The pencil and paper essentially “reconfigure the shape of computational space” so that a once intractable task now becomes tractable. Language is supposed to do the same thing (1997, p. 207). The big question is: how does language reconfigure computational space? And does this reconfiguring of computational spaces really create new kinds of cognitive abilities rather than just offloading information processing for limited cognitive creatures? The most revealing explanation that Clark gives for understanding the kind of things that words do for our cognitive abilities is an analogy he draws with what we might call “scaffolded chimps.” Clark 2002 cites a study in which chimps learned to recognize higher order relations. Chimps were trained (in prior studies) to recognize pairs of objects as either the same or different by associating different tokens with various pairs of objects (e.g. a yellow triangle to designate sameness, such as two shoes, and a red circle to designate difference, such as a shoe and a cup). These chimps who had received training with tokens in prior studies, in contrast with chimps who had not received the prior training, were then able to solve a new task:
correctly identifying instances of higher-order sameness and difference (e.g. a pair consisting of a cup and a shoe and a pair consisting of two shoes would constitute an instance of higher-order difference; two cup/shoe pairs would constitute an instance of higher-order sameness). The ability to learn such higher-order relations was beyond the ken of chimps who had not received prior training relating symbols to first-order sameness and difference relations (Thompson, Oden, and Boyson, 1997, cited in Clark 2002, p. 41).

Clark, following Thompson and Oden 1996, thinks that the symbol-trained chimps’ ability to recognize higher-order relations is due to their retrieving mental representations of the tokens upon confronting instances of the pairs of objects (2002, p. 41). Thus, for example, when looking at a cup/shoe pair, the chimp would retrieve a mental representation (an image, I would suppose) of a red circle, and, alternately, would retrieve a representation of a yellow triangle when looking at a shoe/shoe pair. Clark’s idea is that upon encountering these two pairs, the chimp just compares the two mental images (a red circle and a yellow triangle) to arrive at the correct answer of there being a higher-order difference (ibid., p. 42). This is the kind of thing that Clark means when he talks about “reconfiguring computational space” so that our pattern-completing brains can solve more complex and more abstract problems. It will be instructive to quote Clark himself on this point:

Learning a set of tags and labels (which we all do when we learn a language) is, we may thus speculate, rather closely akin to acquiring a new perceptual modality. For like a perceptual modality, it renders certain features of our world concrete and salient, and allows us to target our thoughts (and learning algorithms) on a new domain of basic objects. This new domain compresses what were previously complex and unruly sensory patterns into simple objects. These simple objects can then be attended to in ways that quickly reveal further (otherwise hidden) patterns, as in the case of relations-between-relations….The
thing I really care about, then, is this process of objectification in which a set of simple tokens stands in for complex features and relations and makes available new, quasi-perceptual, spaces for reasoning (2002, p. 42).

There are two important and interrelated ideas here: language is conceived of as a new domain of objects, and these objects (words) capture or stand in for complex patterns.

What concerns me the most is what Clark might mean by the idea that language “stands in” for complex features and relations. When we use words to think with, we do not actually think about the word itself, but, in some sense, what the word represents. Somehow I use “dam” to think about dams. I do not speak Chinese, so although I could represent the Chinese word for dam (as a certain complex shape or sound) I could not use it to think about dams. How do words allow us to think about the complex patterns they represent? If we are supposed to take our cue from the scaffolded chimps, it seems the answer would be: by forming an association when paired with the complex pattern (i.e. relations of sameness and difference). The chimp is supposed to represent the sameness relation, for example, by associating the yellow triangle with any instance of an identical pair (e.g. shoe/shoe or cup/cup). This allows the chimp to recognize, say, a pair of shoes and a pair of cups as an instance of higher-order sameness simply by ‘seeing’ the image of two yellow triangles in the place of the pair of shoes and the pair of cups. I have no doubt that words can be useful in precisely the way Clark claims the yellow triangle and red circle tokens are useful for the chimps, i.e. reducing a complex or abstract

62 There is, thus, an important difference between the kind of “standing in” that the emulator that Clark and Grush 1999 discuss does (see above) and the kind of “standing in” that words do. The emulator acts as a stand in by essentially acting as a physical duplicate of the signal which the environment would produce. The way to understand representation in this case, it seems, is in terms of (physical or functional) resemblance to the states that would be caused by real world interaction. Words, on the other hand, do not ‘stand in’ for the things that would be caused by real world interactions by resembling them, e.g. “dog” does not resemble the perceptual state caused by dogs.
relationship to a simple matter of perceptual recognition. What I doubt is that this kind of account of the utility of words can explain how words enable us to think about the complex regularities/properties in a way that avoids the problem of abstraction. The basic problem with Clark’s account of the function of words, I claim, is that he doesn’t explain how linguistically-scaffolded thought enables new kinds of thought contents unavailable to nonlinguistic thought. The little that Clark does say about the relationship between linguistic and nonlinguistic thought seem to put him, his own claims to the contrary notwithstanding, within (what he calls) the communicative view of language, according to which language merely expresses thought contents already formulated independently of language, as I will presently show (Clark, 1997, p. 196).

In order to try to think through how words might enable thought about abstract properties, consider a case analogous to the higher-order-relation-recognizing chimps. How might a human recognize that two different looking objects are nevertheless objects of the same kind, for example, how a bag of flour and an egg are each instances of INGREDIENT? Perhaps I can do so because when I see the bag of flour I associate the word “ingredient” with it and when I see the egg I associate the word “ingredient” with it, and then notice that each object is associated with a mental image of (either the shape or sound) the word “ingredient.” This would seem to be analogous to the chimps in that one’s recognition of an abstract property or relation depends on using a word to reduce “abstract” phenomena to a quasi-perceptual problem (in this case imagining two mental images of the word “ingredient”). But now the obvious question is: in virtue of what did I count the flour and the egg as instances of INGREDIENT in the first place? Do I already represent these items as INGREDIENTS or does the word enable me to think
about ingredients as such, which I couldn’t think about without the word? If I already represent them as INGREDIENTS then the word does not capture that complex property, rather, it merely expresses it. On the other hand, how could a word by itself enable me to think about things? There is a Chinese word for dam, but unless I know its meaning it is just meaningless lines on a page.

One might claim that we know that an egg and flour are both ingredients because we have at some point been taught that they are. But this cannot be an ultimately satisfactory answer for how language enables us to recognize instances of a more abstract property. The reason is clear once it is asked: And how does the teacher come to learn that an egg and the flour are both instances of INGREDIENT? One can say that they themselves learn it by being taught by another teacher only at the cost of entering on an infinite regress. In order to avoid that regress one must admit that at some point someone first had the concept INGREDIENT which preceded any learning of it through learning some kind of correlation between words and things. Moreover, there is a basic problem with the account according to which one learns by matching occurrences of words with occurrences of things since many times the things occur without the corresponding word and other times the words occur without the corresponding things. For example, mother may say that she needs to go to the store to buy some ingredients and in this circumstance neither the egg nor the flour are present.

So, then, what is the relationship between words and thoughts for Clark? I think Clark’s answer to these questions is, in the end, expressivist (Gauker, 1994). That this is so is clear since he thinks that words essentially “freeze” and “redescribe” our complex, connectionist-style-represented thoughts in a more abstract format (Clark, 1998, p. 33;
1997, pp. 210, 211). The idea of “freezing” our thoughts in a more abstract form certainly seems to imply preexisting thought contents. The utility of language, for Clark, is that it gives us objects (words), on which we can then focus our attention. This leads to what Clark sees as a new kind of thinking that language makes possible: meta-thinking—thoughts about thoughts. Such second order cognitive abilities are exemplified in “self-evaluation, self criticism, and finely honed remedial responses” (1997, p. 208). The idea, then, is that it is in virtue of the fact that words are objects that we can then take them as objects for further thoughts.

One problem with Clark’s account is that it does not seem that words are required for meta-thinking. What is to keep us from just taking a thought as the object of another thought without having to have a word standing for that thought? For example, what is to keep me from having an image about an image I had an hour ago? Perhaps Clark’s answer here would be that the thoughts which our connectionist-style representational resources enable are much too detailed to be able to serve as objects of further thought (c.f. Clark, 1997, p. 210). The idea here might be, roughly, that although we can recognize certain complex properties in the world if we see them, we cannot recall to mind all the detail we have seen (in the same way that we cannot imagine a chiliagon). For example, although we could not recall all the detail of a face, we could recognize changes in the face if we are looking at it. But if this is so, and if words do enable us to think about these detailed contents, the question remains: How are the words related to the detailed thought contents? Again, when I’m thinking about dams using the word “dam” I’m not thinking about the particular shape or sound of that word, I’m thinking about dams. Clark thinks that language somehow captures our thoughts in a more
abstract way that does not simply recapitulate them (1997, p. 210), but he has given us no account of the relationship between the word and the more abstract content words purportedly capture other than some vague suggestions such as language “freezing” thoughts or “redescribing” them—language that would seem to suggest some sort of abstractive process.

In fact, Clark has subscribed to the idea of an abstractive mechanism (Clark and Karmiloff-Smith, 1993). In particular, he accepts the idea of “skeletonization” according to which the least relevant hidden units in a neural network are discarded in building more simplified, abstract representations of that network’s ‘knowledge’ (ibid., pp. 507-508). In addition, he accepts some version of the neural control hypotheses (1997, pp. 136-141), which all fall under the category of what, in chapter 1, I called the vector space theory of concepts, and thus face the problem of abstraction. Nevertheless, Clark owes us an explanation of how words are related to the more abstract contents that he claims they make possible. And it seems like the dilemma he will face here is that either words will express the more abstract contents already encoded in some nonlinguistic format or they will themselves create such contents. The latter option is false: there is nothing intrinsic to the image of a word that makes it fit to create a particular content, whereas the former faces the problem of abstraction. Words may yet be able to create more abstract contents, but this will not be due to any intrinsic properties of the word qua image. In chapter six I will present an account of how words could possibly create more abstract contents.

63 In chapter 2 I criticized a particular version of the vector space theory of concepts forwarded by Larry Barsalou (1999, 2005). The neural control hypotheses Clark (1997, pp. 136-141) discusses, e.g. Damasio’s convergence zones hypothesis, would face the same problems as Barsalou.
As I see it, the basic problem for Clark comes down to this. Although he makes good enough sense of how words (or mental images) could reconfigure problems to make them solvable—as in the case of the scaffolded chimps, he fails to explain the relationship between the word and the thought contents that the word makes possible. For example, it seems unlikely that a chimpanzee could think the thought NO PLANT IS ALSO AN ANIMAL, and perhaps this is so because a chimpanzee has no way to think about superordinate relationships such as that a dog and a snake are both animals. And perhaps, as Clark suggests, they cannot think about these higher-order relationships because they do not have word-objects to help them think about such higher-order relationships. Perhaps the scaffolded chimps could solve problems which compared various plants and animals if they had been trained with two different tokens corresponding to plants and animals. It would not follow, however, that if chimps could solve such a problem that they could think thoughts like NO PLANT IS AN ANIMAL. And the reason is that solving a task using words as a scaffold is not yet to explain how words might create the ability to think about animals or to think about plants. In particular, it suggests no account of how the word “plant” and the word “animal” are related to the concept ANIMAL and the concept PLANT, which figure into judgments such as NO PLANT IS AN ANIMAL.

Part of the issue seems to be the way Clark has posed the problem in the first place. He prefers not to answer questions such as whether or not we think in language but, rather, to ask how language can help us solve problems (1997, p. 200). But we are never going to understand the more distinctive ways words augment cognition (and help us solve problems) without understanding the semantic properties of words—how words
are related to the kinds of thoughts we can think. We cannot understand how symbols help us solve problems until we understand how they function as symbols. But Clark is frustratingly vacuous on this point.\textsuperscript{64}

\textsuperscript{64} Consider Clark’s discussion of our representation of the number 98: “So is there an internal representation of, let’s say, ‘98’? I think the answer that now emerges is ‘No’! What we have are genuine internal representations of the word ‘ninety-eight’ (or perhaps the phonetics of the word) and the numeral ‘98’, and genuine internal representations of rough quantity and relative location in an array” (Clark, 2002, p. 89). Our “genuine internal representations,” according to Clark, are analogous to a kind of subitizing ability according to which we can judge differences in magnitude (e.g. between 3 vs. 9 but not between 6 vs. 7, say) without representing discrete, precise quantities. But, again, the question remains: how does the numeral “98” allow us to think about that quantity? We aren’t merely thinking about the marks on the page or the phonetics of the word when we think “98”.
**Daniel Dennett: Cognitive offloading**

Dennett has for many years (e.g. Dennett 1971) defended a view according to which intelligent behavior is ultimately to be explained in terms of “dumb,” mechanisms, whose specialized roles have evolved because they have enabled adaptive behaviors. Much of Dennett’s corpus on intentionality and consciousness can be seen, at root, as carrying out this project: a project which carries on the tradition of exorcizing the ghost from the machine—or in Dennett’s terms, explaining the mind with cranes rather than skyhooks (1995, p. 73). Famously, Dennett has defended the idea that our attributions of intentionality to a system are warranted to the extent that such attributions enable us to explain and predict the behaviors of a system whose internal workings we fail to understand. Although we can explain the workings of a typewriter by understanding its design, and can explain why the typewriter isn’t functioning by invoking physical laws (e.g. the typewriter won’t work because its not plugged in), some systems, such as a chess playing computer would be very difficult to explain using either the “design stance” or the “physical stance” (Dennett 1971). Rather, in such a case we would adopt the “intentional stance” which involves attributing information and goals (the analogues of beliefs and desires) to the system and then working out the most reasonable action for the system to take in light of these attributions. Taking the intentional stance is instrumental and pragmatic: it is due to our explanatory and prediction purposes, and attributing such goal and information states may not give us any guidance about the system’s design: “intentional theory is vacuous for psychology” (Dennett 1981, p. ) Thus, our attribution of intentional states to a system is epistemic and pragmatic, not ontological: it does not mean that the system does not reduce, ultimately, to simple,
nonintentional mechanisms. Indeed, for Dennett intentional explanations must ultimately be replaceable by nonintentionally-specified mechanisms.

Unlike his teacher Gilbert Ryle, Dennett is not opposed to positing internal representations (1987, p. 218), for they need be no more problematic than the internal representations of a chess playing machine or a calculator. What one must avoid is positing representations which cannot ultimately be explained in terms of the functional role they play in a mindless input-output system: in explaining human cognition there can be no left over mind stuff at the end of our analysis. However, there is no a priori guarantee that we will be able to functionally decompose a system as complex as the human brain into discrete elements that straightforwardly correspond to a semantic interpretation (1987, p. 224). The reason for this follows from the fact that nature, unlike human engineers, is a short-sighted tinkerer which sometimes co-opts an already functioning part into another functional role, thus creating parts which have multiple functions (1995, p. 216; 1991, p. 273; 2001, p. 141). And if, as Dennett claims, a semantic interpretation must correspond to the global role that the part plays in the behavior of the system, then given the multiple functionality, there will be no simple semantic interpretation of the parts.

Dennett’s belief, then, is that we can explain both content and consciousness using only complex assemblies of mindless automata (cranes) performing their various functions in concert. Our intentionality is not sui generis and doesn’t differ fundamentally from that of a chimpanzee or an intelligent robot (1995, pp. 422-426). How, then, are we to explain what seem to be large gaps between an intelligent robot and a human? Characteristically human thought, Dennett thinks, has quite a lot to do with the
fact that we have language (1995, chapter 13; 1996; 1991, chapters 7, 8; 2000; 2001). Language is a crane which non-mysteriously explains some of the apparent gaps between cognitive abilities of humans and those of a chimpanzee or an (imagined) intelligent robot. How might language do this?

Dennett frames the issue of the role of language in cognitive development in terms of a progression of different types of problem solvers (1996, chapter 4; 1995, chapter 13). The most evolutionarily primitive problem solvers, Darwinian creatures, approached the world with a stock of hard-wired attributes. This set of hard-wired attributes (the phenotype) enabled such creatures to solve certain kinds of problems but, lacking phenotypic plasticity, they could not learn new ways to solve problems. The next type of problem solver, the Skinnerian creature, was able to adapt its behaviors to novel and changing circumstances, but had to “blindly” generate new behaviors, some of which would have been maladaptive—in the worst case leading to the creature’s death. In some cases, however, the behavior would be successful and would be reinforced because of its success. It is with the next kind of problem solver, the Popperian creature, that representation emerges. Unlike Skinnerian creatures, Popperian creatures have an “inner environment” which contains information about the regularities in the world (Dennett 1996, p. 88; c.f. Cummins 1996), the advantage of which would be that the Popperian creature does not have to risk its own welfare in trying to figure out which behaviors would be adaptive in a particular circumstance.\footnote{Dennett calls these creatures “Popperian” after Sir Karl Popper who noted that such foresight “permits our hypotheses to die in our stead” (Dennett 1995, p. 375).} Popperian creatures are thus able to utilize information in a way that Skinnerian creatures cannot; but having information isn’t an all or nothing affair: there is always more and better information to be gained.
Gregorian creatures are distinguished from Popperian creatures because of their ability to import designed artifacts—of which words are the most important—into their inner environments. These artifacts enable them to be even better problem solvers by giving them both more and better information as well as a new ability to access and manipulate that information. These two aspects (increased information and access to that information) constitute the core of Dennett’s view about the cognitive functions of language.

To take the former first, there are several ways that artifacts can give us more information with which we can better carry out tasks and solve problems. First, there is a sense in which artifacts contain information which we can use without having to actually represent. For example, an artifact like a pair of scissors not only takes intelligence to make but also confers intelligence on its user by increasing one’s ability to arrive at better solutions to the tasks they need to perform (Dennett 1996, p. 99). Because someone at some point solved a certain problem by designing an artifact, we can benefit from the information that was used in building that artifact. We can benefit simply by knowing how to use the artifact, whether or not we know how it works or what particular constraints informed its design. In this sense, we benefit from the information used in the design without having to represent that information (e.g. come up with the particular solution) ourselves.

Secondly, artifacts can help us offload some of the cognitive work of problem solving, thus allowing us to solve more complex problems. For example, a computer spreadsheet program may enable me to solve various tasks that would be difficult, if not intractable, to solve without such a program. The reason for this is simply that we do not
have unlimited memories nor the ability to perform complex operations on data structures. A simpler example of cognitive offloading would be simply using a pencil and paper to solve a long division problem. Here we have both an algorithmic technique (long division) and the resources (pencil and paper) which essentially allows us to bring our limited cognitive capacities to bear on simpler, more tractable tasks (e.g. recognizing that $7 \times 7 = 49$).

How, then, does language fulfill these roles? It is one thing to explain how artifacts such as a pencil and paper and the technique of long division allow us to offload some of the cognitive load onto various artifacts—that is, by utilizing a technique that breaks a complex problem into simpler parts, and storing the intermediate results on the paper rather than in memory—but it is not initially clear in what way language is analogous to artifacts such as these. Dennett’s explanation of how language-as-artifact increases cognitive ability is very similar to Clark’s: words function as found objects (1996, p. 150; 2000, p. 24; 1993, p. 546) which become linked to preexisting representations in the brain. As objects, however, words can be thought about and manipulated in a way that the implicit, procedural representation of nonlinguistic creatures cannot (1996, p. 151). For Dennett there are several benefits that accrue from word-objects: they allow us to think thoughts about identity (e.g. that penny is the same penny I brought from New York), to call to mind solutions to problem situations that are not present, and to metarepresent—to think about and evaluate our own thoughts. But the question which remains to be answered is how words qua objects create these new

---

66 The direction of influence here is not clear. Dennett (2000, pp. 17, 24) cites Clark 1997 as influencing his thinking on the nature metarepresentation.
cognitive abilities? And how are the words related to the preexisting nonlinguistic representations in the brain?

Dennett’s answer to the latter question, like Clark’s, seems to be committed to an account according to which words merely express contents already represented in the brain. More precisely, words as sounds become familiar in their own right, without being understood, and then become attached to certain meanings or experiences. Consider a few key passages from Dennett:

A word can become familiar even without being understood. And it is these anchors of familiarity that could give a label an independent identity within a system. Without such independence, labels are invisible. For a word to serve as a useful, manipulable label in the refinement of the resources of the brain, it must be a ready enhancer of sought-for associations that are already to some extent laid down in the system (1996, p. 150, my emphasis).

The habit of semi-understood self-commentary could, I am suggesting, be the origin of the practice of deliberate labeling, in words (or scribble or other private neologisms), which in turn could lead to a still more efficient practice: dropping all or most of the auditory and articulatory associations and just relying on the rest of the associations (and association possibilities) to do the anchoring. The child, I suggest, can abandon out-loud mouthings and create private, unvoiced neologisms as labels for features of its own activities (ibid., my emphasis).

*Once we have created labels and acquired the habit of attaching them to the experienced circumstances, we have created a new class of objects that can themselves become the objects of all the pattern recognition machinery, association-building machinery, and so forth. Like the scientists lingering retrospectively over an unhurried examination of the photographs they took in the heat of experimental battle, we can reflect on whatever patterns there are to be discerned in the various labeled exhibits we dredge out of memory* (ibid., my emphases).

In these passages Dennett seems to accept that words are somehow related to preexisting meanings or contents (“associations,” “features of its own activities,” “experienced circumstances,” “exhibits we dredge out of memory”). Thus, when it comes to nonlinguistic creatures, the issue is not that the creature doesn’t have the procedural
know-how which such brain representations make possible, but that without words there is no way to isolate and access that procedural know-how:

A polar bear is competent vis-à-vis snow in many ways that a lion is not, so in one sense a polar bear has a concept that the lion lacks—a concept of snow. But no languageless mammal can have the concept of snow in the way we can, because a languageless mammal has no way of considering snow “in general” or “in itself.” This is not for the trivial reason that it doesn’t have a (natural language) word for snow but because without a natural language it has no talent for wrestling concepts from their interwoven connectionist nests and manipulating them. We can speak of the polar bear’s implicit or procedural knowledge of snow (the polar bear’s snow-how), and we can even investigate, empirically, the extension of the polar bear’s embedded snow-concept, but then bear in mind that this is not a wieldable concept for the polar bear (ibid., p. 159).

Here, then, is Dennett’s idea in brief: the value of language is to be seen within the context of creatures which generate solutions to problems. Creatures which represent are able to make better than chance first responses to a problem, and creatures with language are able to make even better first moves than nonlinguistic creatures because of the more flexible representational capacities which language makes possible. The increased flexibility of representation is due to the practice of cognitive offloading, and language fulfills this role by representing complex associations in another, simpler format that we can access and manipulate, namely words. Moreover, since words are objects for us, we can use them to think about situations in which the things the words represent are not sensibly present (1996, pp. 157, 159). Such re-representation also makes possible a new ability hitherto unavailable to nonlinguistic creatures: meta-thinking or what Dennett calls “florid thought,” which is the ability to reflect on and consider our own thoughts. Again, this is so, according to Dennett, because words are objects for us and as such can be the objects of further thoughts. Words (as patterns of sound or marks on a page) are
literally objects and they come to be associated with certain kinds of experiences and patterns of association (such as might be captured in a connectionist network).

Thus Dennett shares with Clark both the idea of language as a tool for offloading cognitive work and the idea that language creates metacognition. Dennett gives no more of an idea of how language accomplishes these feats than Clark does. Thus, it seems that Dennett faces the same dilemma that I posed for Clark: either words express already laid down meanings, or they create more abstract meanings. If they express the very meanings that are already laid down, then language is not playing any more of a cognitive role than something like a necessary crutch for memory-limited systems—a position which is consistent with Fodor 1975. On such a view, all our concepts precede language, but language can serve as a useful mnemonic for the complex formula of the language of thought (Fodor 1975, p. 84). If, on the other hand, words actually create more abstract meanings then the problem is to explain how a word qua sounds or marks on a page can carry meaning. How do words themselves create new semantic properties (e.g. more abstract meanings)?

As I noted above, it seems that Dennett at times thinks of words as becoming attached to meanings which the words then express. On the other hand, he also thinks that communication puts a pressure on people to formulate their thoughts and that such formulation always creates a confabulation since the rough-grainedness of words can never capture the fine-grainedness of what is actually going on in the brain (Dennett 1996, pp. 127-128). Moreover, Dennett also sometimes says things that seem to be clearly anti-expressivist, such as when he claims that talking isn’t expressing some

---

67 See above, p. 15.
already expressed thought, but that the very process of talking creates the content of the thought (1991, p. 315). These tensions can be reconciled, I think. The confabulation/anti-expressivist point could just be that we always fall short of capturing in our speech the complexities which actually give rise to our behaviors. And this is consistent with the idea that words get their meanings by becoming associated with certain activities and experiences, since it could just be that the meanings we create are never fine-grained enough to capture the complex, interconnected, and more primitive control systems which are actually responsible for our behavior. Thus, words would indeed express the meanings with which they are associated, but these meanings would be confabulated.68

Thus, if Dennett thinks that words express concepts whose contents are established independently of those words, then the question is: where do the semantic properties of concepts come from? It seems the answer is that they come from the “already laid down associations” and “experienced circumstances” stored in memory. And if that is the case, then Dennett too faces the problem of abstraction. Dennett may be right that concepts qua procedural knowledge (in nonlinguistic creatures) cannot be extracted from their “connectionist nests” but he is not right that words can somehow extract these concepts and make them explicit and wieldable. The desire to somehow connect the unique semantic properties which words somehow create to underlying procedural, connectionist representations makes it pretty clear that Dennett is working within (what in chapter 6 I will call) the continuity framework. But, as we have seen, attempting to connect prelinguistic and linguistic thought in this way makes it difficult to

68 Of course, we would still need an account of the where the confabulated content comes from and how it is that different individuals can share this content.
see how language makes possible a new kind of thinking, unavailable to nonlinguistic creatures. It is interesting to note that although both Dennett and Clark want to understand the cognitive functions of language in terms of problem solving, neither one countenances the idea that the most distinctive contribution of words may be the role they play in interpersonal, communal interactions. In chapter six I will suggest an account of the significance of language along these lines. Also, in the last chapter (chapter four) I suggested an account of the significance of language in coming to understand false belief according to which the interpersonal role of language is paramount.
José Luis Bermúdez: Language and Personal-level Inference

Bermúdez 2003 argues that we must engage in a limited form of propositional attitude psychology in order to explain certain behaviors of nonlinguistic creatures. In order to get to this conclusion, he first lays out a minimalist conception of nonlinguistic thought and then argues that certain kinds of thoughtful behaviors, such as tool use, cannot be explained on the minimalist construal of thought. Bermúdez then goes on to argue that the kind of thought available to nonlinguistic creatures cannot account for the range of behaviors collectively referred to as second order cognitive dynamics—i.e. the “conscious regulation and policing of one’s own thoughts” (p. 159). Bermúdez cites approvingly Andy Clark’s thesis that language makes second order cognitive dynamics possible (see above) and sees himself as providing a more precise account of this basic idea. My aim in the following is to flesh out some of the details of Bermúdez’ basic two step argument and raise some objections at certain key points. In particular, I will question his characterization of the minimalist construal of thought as limited to the here-and-now. I will then suggest a way of construing minimalism so as to account for certain kinds of cases that Bermúdez thinks require positing propositional attitudes. Finally, I will suggest that Bermúdez sees linguistic and nonlinguistic thought as largely continuous in that both can be characterized in terms of the same kind of representational content.

Minimalism

Bermúdez elucidates the minimalist construal of nonlinguistic thought by developing the intuitive distinction between knowing-that and knowing-how (Ryle, 1949). Whereas
knowledge-that requires specification using a proposition (e.g. Bill knows that the house is on fire), know-how does not obviously require specification with a proposition (e.g. Bill knows how to swing a golf club). In this latter case, Bill need not be entertaining any propositions about golf clubs. Rather, he just needs to have the complex bodily skill of swinging the golf club. There are many bodily and imagistic abilities like this (e.g. climbing a rock wall, riding a bicycle, playing foosball, imagining how to put something back together, imagining if some object will fit in a certain space, etc.). Minimalism, according to Bermúdez, is committed to explaining all nonlinguistic thought as a version of nonpropositional know-how.

Bermúdez takes perception as the obvious starting point for thinking about the characteristics of minimalist thought. Following philosophers like Dummett, Cussins and Campbell, Bermúdez thinks that minimalist thought is nonpropositional in the sense that it is unstructured, perceptually vehicled, pragmatic and tied to the here-and-now (2003, p. 45). Taking perception as our model, Bermúdez claims that minimalist thought contents are unstructured in the sense that their contents cannot figure into inferences (in a similar way to how pictures are not representations that can feature in inferences). Furthermore, they are indeterminate in that there could be many different propositions that could equally well be said to capture the content of perception (just as many different sentences could describe the same picture). Minimalist thought is essentially pragmatic in that it is tied to the possibilities for action in the creature’s environment (ibid., p. 62). Bermúdez also thinks that minimalist thoughts cannot extend beyond the possibilities for action that a particular context affords. His reason for thinking this is so seems to just be that he adopts Dummett’s account of protothoughts, according to which the thinking cannot be
detached from present (ibid., p. 42). Since the twin characteristics of being unstructured and tied to the present turn out to be key for Bermúdez’s argument for the insufficiency of minimalism to explain all nonlinguistic thought, it will be important to further investigate the rationale for including them in the first place.

Bermúdez seems to think that minimalist thought is tied to the here-and-now as the result of being action oriented. Several times he approvingly quotes Dummett’s characterization of protot thoughts as “the superposition of spatial images on spatial perceptions” (2003, pp. 43, 44, 47). He also thinks that another way of thinking about minimalist thought is in terms of J. J. Gibson’s notion of affordances according to which the environment is perceived in terms of possibilities for action rather than in neutral terms (Gibson, 1979). Of course, if such characterizations were correct, then minimalist thought would indeed be tied to the here-and-now since a necessary component of thought would be current perceptions. But inasmuch as Bermúdez is attempting to flesh out the intuitive notion of knowing how, any such restriction of minimalist thought to the here and now should give us pause. To see why, consider one of the examples of thinking how that Bermúdez gives: imagistic reasoning. Bermúdez’s example is that of “calculating whether a wardrobe will fit into the space between a chest-of-drawers and the bed”—a feat which he claims is “carried out by exercising our visual imagination” (2003, p. 36). The question I want to raise is: Is visual imagination necessarily tied to the here-and-now, or can one manipulate visual images of situations which are not currently present? No doubt many times such a use of images is tied to the here and now, but need it be? Bermúdez seems to tacitly assume so, and it is this assumption which plays a key
role in some of his arguments for the insufficiency of minimalism to explain all
intelligent actions of nonlinguistic creatures.

However, it doesn’t seem to me that there is any good reason for assuming that
visual imagination cannot be exercised other than by manipulating objects (in
imagination) which are sensibly present. After all, it seems that we can use images to
think about and solve problems regarding situations which we aren’t currently
confronting. I can imagine alternate ways of getting home in case normal routes are
blocked, and I can search through the kitchen drawer to find something which will allow
me to turn that awkward screw on my computer upstairs. Likewise, I might choose to
take the left fork of the trail rather than the right, having gone down the right fork trail in
my imagination and encountering that impassible mud hole. These kinds of things seem
to both involve a kind of visual imagination and be detached from the here-and-now since
that which I am reasoning about is not present to my senses in any of these cases.
Likewise, in Bermúdez’s example of rearranging furniture it may be easier to do while
standing in the room with the furniture, but it does not seem impossible to do if I were
not standing in the room. This makes me think that having the objects about which we’re
thinking present may indeed be easier (for our limited memory and processing capacities)
but is in no sense required.69 I will come back to this point in discussing Bermúdez’s
arguments for the insufficiency of minimalism.

The other key characteristic of minimalist thought is that it is unstructured.
Basically, this means that minimalist thought lacks the object-property structure of
propositional thought—alternately, the subject-predicate structure of judgments (c.f.

69 It also seems clear that some people’s visual imagination is better than others. At one
extreme would be people like Luria’s “S” (Luria 1987).
2003, pp. 59, 89). That minimalist thought lacks such structure would seem clear if it is imagistic or perceptual since such contents cannot be combined compositionally (with either themselves or other propositions) and hence cannot feature in inferences. I will not take issue with this characteristic of minimalism since my own characterization of minimalism developed in Chapters 1 and 2 likewise accepts that nonconceptual representation (i.e., representation in a similarity space) is logically unstructured (no discrete subject-predicate represented) and noncompositional (because there are no conceptual components to combine and recombine).

**Psychological explanation according to minimalism**

The essence of a psychological explanation, Bermúdez thinks, is that it invokes how a creature represents its environment in order to explain its behavior (2003, p. 8). And this, in turn, is only necessary when behavior cannot be predicted by the mere “registering” of an appropriate stimulus (ibid.). Thus, something like classical conditioning or innate releasing mechanisms are not psychological explanations because there is an invariant relation between the stimulus and the response that makes appeal to representational states unnecessary. In contrast, any behaviors which are either flexible or plastic in response to the environment are the kinds of behaviors which are prima facie candidates for psychological explanation. The kind of states minimalism invokes are nonpropositionally construed and consist of a perceptual state and goal desire, both of which are unstructured and thus compatible with the minimalist notion of representation. As noted above, Bermúdez thinks (unjustifiably, I have argued) that the representational states postulated by minimalism must be tied to the here-and-now, and so in some sense
must be tied to perception. Since it is also possible for the creature to act on mediately perceived instrumental properties (such as when one is alerted to the presence of a predator by the behavior of the creature’s own conspecifics), the content of the perception can include things that are not directly perceived (ibid., pp. 52-53).

Desires must be accounted for in terms of what Bermúdez calls “goal desires.” Goal desires, in contrast with situation desires, represent the goal in the most general terms. For example, the goal of attaining food in contrast with the goal of eating a specific food from a specific place. Situation desires would require specification with a “that—” clause, as in desiring that I eat Indian food at Krishna. Bermúdez cites an example of the marsh tit which caches food in hundreds of different locations and then returns later to retrieve the food. In this case he thinks it would be incorrect to say that the bird is being driven to obtain food from a certain location. Rather, it is being driven by the more general desire for food tout court (ibid., p. 48). Goal desires, Bermúdez thinks, are not structured and so are kosher for the minimalist. They are not structured because, as noted, they do not require specification with a “that—” clause. All one needs to do in order to cite a goal desire in an explanation is to work backwards from the attained goal to the goal desire (ibid., p. 49). Bermúdez notes that one need not really even specify the content of the perceptual state since we can expect that the creature is suitably sensitive to the possibilities for action that the environment affords (ibid.). Thus all one really needs to do in giving an minimalist explanation is cite the goal desire.

One thing that seems strange to me about Bermúdez’s account of “goal desire” is that—contra his own requirements of psychological explanation—it does not seem to require representation at all. Or at least it only requires representation in a very
attenuated (bordering on vacuous) sense. According to him we cite a current perception and a goal desire, but neither one of these seem to require mental representation. Take the desire for food, for example. I may have the goal desire for food if my stomach is growling, but need I have a mental representation corresponding to the desire for food? It seems not. Rather, all I need are the unpleasant sensations in my stomach that drive me to find food, and these do not seem like good candidates for mental representations. Of course, once I have those stomach sensations I may need to do some thinking to figure out how to go about getting food, and this thinking may plausibly involve mental representations, but it doesn’t seem that the desire for food (i.e. hunger) itself need be understood in terms of mental representations. Situation desires, of course, would need to be represented but these are not allowed on the minimalist account. We need to invoke psychological explanations when the stimulus-response relationship is not invariant, and we may indeed have to invoke the goal desire for food to explain certain behaviors. My point is just that its not obvious that the desire for food need itself be understood as any kind of mental representation—whether propositional or nonpropositional.

*What minimalism leaves out*

According to Bermúdez,

we will have an argument against the minimalist interpretation of nonlinguistic thought if we provide examples of behaviors in nonlinguistic creatures that do not seem to be interpretable at the level of perceptual content, whether that content is understood in terms of affordances, or some other form of noninferential mediate perception (which might, for example, involve the exercise of visual imagination in reconfiguring the elements of the perceived environment) (2003, p. 53).
And there are such examples; whether or not they show that we must engage in a form of propositional attitude psychology is another matter. One example Bermúdez gives is that of tool construction by wild chimpanzees:

Wild chimpanzees, for example, make two different types of wands for dipping into ant and termite nests from different types of branches. Wands for dipping into ant swarms are made by stripping the side leaves and leafy stem from a stick several feet long. The wands constructed for dipping into termite nests, on the other hand, are made from vines or more flexible twigs and are considerably shorter. They also have a bitten end, unlike the ant wands. These tools are extremely specialized and, as Byrne notes (1995, 97), the wands are often constructed some time in advance and a considerable distance away from the place where they are going to be used (Bermúdez 2003, p. 55).

The behaviors of these wild chimpanzees cannot be explained within the resources of minimalism according to Bermúdez because they involve making inferences about things that are not even mediately present (pp. 54-55). In particular, Bermúdez seems to think that the goal of extracting termites is not mediately perceptible. Thus, he concludes that they must be representing the relationship between a current action and a future goal, and seeing the action as leading to (in some inferential sense) the future goal (ibid.). If such representation were required, then I agree: it could not be understood on the minimalist construal of thought. What I have questioned, however, is whether it is correct to limit minimalism to the immediate and mediate perception. If minimalism were not limited in this way, then the minimalist could help herself to an “imagistic reasoning” explanation of the chimpanzee’s behavior, obviating the need to invoke any kind of propositional attitude psychology. Supposing that the chimps have had a good amount of experience attempting to extract termites from their nests, why couldn’t they imagine (by, so to speak, putting themselves in the situation of extracting termites) what the tool needs to look like in order to successfully extract the termites? Just as when I am in the kitchen
rummaging around through the drawers to find an adequate tool while holding in mind the visual and kinesthetic images related to the problem task, so the chimpanzee could use imagery to adapt the branch to the problem task. Such “imagistic reasoning” stands in contrast to representing the inferential relationships between actions and goals.

Bermúdez also thinks that minimalism also fails to account for how animals can learn from experience. Since it is clear that animals do learn from experience, then minimalism cannot be the whole story. His argument for this depends on establishing: 1) that minimalist representation is unstructured and 2) that some kinds of learning require structured representations. The first follows from Bermúdez’s characterization of minimalist thought as perceptual and imagistic. The argument for the second goes like this: Being able to represent contingencies between objects in one’s environment requires that one’s representations be “structured so as to have separable components that can reappear in, and be extrapolated from, a range of further representations of the environment” (2003, p. 57; c.f. 1995, p. 350). Although the minimalist can accommodate the tracking of features in the environment, he cannot track regularities between objects because doing so implies a kind of representation which is both detached from the here-and-now and not constrained by the requirements of immediate action (2003, p. 59), neither of which are (on Bermúdez’s characterization) within the purview of minimalist thought. Bermúdez explains the difference in terms of Strawson’s distinction between feature-placing level and the particular-involving level:

At the feature-placing level of experience there is no distinction to be drawn between one instantiation of, say, foodness and another. A creature at this level inhabits a world composed of different kinds of stuff. These different kinds of stuff can be encountered at different places, at different times, and in different combinations. Some of these encounters and combinations are predictable, and the creature best able to make these predictions will flourish. But the regularities
on which these predictions are based are simply conjunctions of features—associations of food with a particular perceptual Gestalt associated with a certain combination of leaves and branches, for example. In contrast, the creature that operates at the particular-involving level is able to detect regularities of a completely different order—regularities that govern the behavior of persisting bodies. Many of these regularities will be causal, governing the interactions between persisting bodies. Others will be kinematic, governing the possible movements and behavior of any given body (ibid., pp. 58-59).

Being able to represent at the particular-involving level requires a kind of structure that approaches subject-predicate judgments (ibid., p. 59). Minimalist representation, in contrast, does not parse an array into distinct objects with properties. Thus, minimalist representation could account for how a creature could perceive general similarities (e.g. as when a creature perceives a tree as affording shelter because it recognizes its similarity to other things which have afforded shelter in the past) but could not account for behaviors which imply that the creature is representing relationships between trees and shadows or trees and birds or between its own actions and a desired outcome.

Bermúdez thinks that there are behaviors which imply that the creature is representing causal relationships between objects and that this ultimately dooms minimalism. It is characteristic of what Bermúdez calls level 2 rationality that genuine decision making occurs: the creature is representing the contingencies between an action and its outcome and then choosing an action on the basis of this contingency (2003, pp. 124, 125, 138). Again, Bermúdez cites the tool-making chimpanzees as evidence that such contingencies are represented (ibid., p. 126). In general, we can say that a creature is representing contingencies if the creature is sensitive “to the consequences of its behavior and to the contingency (or lack of it) between behavior and desired outcome” (ibid., p. 139). So, for example, a case in which we could conclude that the creature is
not representing the contingency would be one in which the creature’s behavior did not ever adjust to a contingency which no longer holds. For example:

A classic experiment has been done by Hershberger (1986), who set up a graphic illustration of such a reversed contingency for chicks. In the experimental setup their food source retreated from them at twice the rate they walked toward it but advanced toward them at twice the rate they walked away from it. Even after 100 trials, the chicks only succeeded in obtaining food 30 percent of the time, which has been taken clearly to indicate that they were failing to represent the two relevant contingencies (that walking backwards causes the food to advance and walking forward causes the food to retreat) (2003, p. 125).

The question is: How are we to understand the kind of practical reasoning that goes on at the nonlinguistic level? It is clear that we cannot apply an inference-based model of practical rationality to nonlinguistic creatures, since it is clear that they themselves could not characterize the logical relationship between their beliefs and desires and the action performed. As Bermúdez notes, this is not the least of which because we have no account of what inference might look like with anything other than linguistic vehicles (ibid., p. 111). Thus, the form of reasoning available to nonlinguistic creatures, Bermúdez concludes, must not require that the nonlinguistic creatures themselves perform any inferences. So what might nonlinguistic reasoning look like?

Bermúdez develops an account of two different forms of nonlinguistic reasoning: protonegation and protocausation. For reasons that will be laid out (and disputed) in the next subsection, Bermúdez thinks that nonlinguistic creatures cannot perform logical operations on whole thoughts. For example, a nonlinguistic creature could not employ any of the truth-functional logical connectives in reasoning since these are operations defined over whole thoughts (propositions). Thus, his account of protonegation and protocausation cannot make any reference to any of the logical operators (hence the prefix “proto”).
Bermúdez thinks that creatures are capable of a kind of disjunctive syllogism: A or B, not-A, therefore B. This is equivalent to the material conditional: If not-A then B. But since sentential negation is defined over whole thoughts, Bermúdez develops a version of predicate negation which is not an operation on whole thoughts. Suppose a creature were reasoning such: If the gazelle is at the watering hole then the lion is not at the watering hole. Bermúdez’s idea is that we can understand this in terms of the contrary concepts PRESENCE and ABSENCE, thus without having to posit sentential negation. So the animal is to be understood as thinking: If the gazelle is present at the watering hole then the lion is absent from the watering hole. Thus, “the gazelle is present” would be the contrary of “the gazelle is absent” and yet not be constructed from it. If a creature could use such contrary concepts then it could engage in a kind of proto-modus tollens: If the gazelle is present at the watering hole then the lion is absent, but the lion is present, so the gazelle is absent. Such an argument would be semantically valid even if not formally valid.

Bermúdez suggests that the proto-conditional should be understood in terms of a kind of causal reasoning. Unlike the conditional, which holds between propositions, causal reasoning would be about states of affairs and so would be kosher on Bermúdez’s prohibition of operators on whole thoughts.

If negation at the nonlinguistic level is understood in terms of the mastery of pairs of contrary predicates…and if we view the relevant conditionals as protocausal conditionals holding between states of affairs rather than between complete thoughts (e.g. between modes of presentation of those states of affairs), then we have all we need for nonlinguistic analogues of our three basic forms of inference. Modus ponens can be understood straightforwardly in terms of a causal conditional together with an understanding, which may take the form of a perception or a memory, that the antecedent holds. The consequent will be straightforwardly detached. We can view modus tollens in terms of the combination of a causal conditional with the protonegation of the consequent of
that causal conditional resulting in the detachment of the protonegation of the antecedent (2003, p. 147).

But of course one must somehow represent the causal relationship if one is to utilize it in reasoning. Bermúdez thinks that causal conditionals are just representations of regularities—either between one’s own actions and the consequences or between states of affairs in the environment (as the gazelle/lion case). The question is: Can a creature represent a causal relationship in such a way that it can “detach” either the antecedent or consequent? At the very least, there must be separable components in order for there to be any detaching of the antecedent or consequent. But this raises the question of how they are attached in the first place, since Bermúdez will not allow logical operations on whole thoughts. Surely “the gazelle is present” is a complete thought, as is “the lion is absent.” Thus, it seems that the representation of the regularity must be structured in such a way that it could be utilized in the protoinferences Bermúdez is recommending. But that makes it seem that there is some sort of operation being performed on two distinct, complete thoughts.

*What nonlinguistic creatures can’t do*

Bermúdez thinks that language brings with it a kind of thought that is unique to language users: second-order cognitive dynamics, e.g. self-criticism and self-evaluation, both of which involve evaluating inferential relations between thoughts (ibid., p. 159). He sees himself as fleshing out Andy Clark’s idea and giving an argument for why such thinking is impossible for nonlinguistic creatures. His argument is as follows:

1. The vehicles of second order thoughts must be either “personal-level” or “subpersonal-level.”
2. Second order cognitive dynamics requires “conscious regulation and policing of one’s own thoughts” and we do not have “direct and conscious access to subpersonal states” (ibid., p. 159). Therefore, the vehicles cannot be subpersonal.

3. So the vehicles of second order thoughts must be personal-level.

4. The personal level vehicles must be either language or images (ibid., p. 160).

5. But images cannot be used to carry out the inferences characteristic of second order cognitive dynamics because they do not have the right structure (ibid., p. 161)

6. Therefore, language is the only personal level vehicle with the right structure to be able to carry out the inferences characteristic of second-order cognitive dynamics.

An example of second order cognitive dynamics would be revising one’s beliefs as the result of reflecting on the relations of evidential support for beliefs (what Bermúdez terms “reflective doxastic modification”). In such cases, someone might come to reject a belief not by encountering some new piece of evidence that conflicts with one’s beliefs but just by noticing that there is little or no evidence for that belief. For example, I might come to reject my belief in God simply because I realize that the hypothesis lacks warrant, despite the fact that nothing ever contradicts it. It is this kind of internalist notion of justification that serves as a paradigm case of the kind of second order cognitive dynamics which Bermúdez thinks is unavailable to nonlinguistic creatures.

Bermúdez may be correct to say that personal-level inferences can only be carried out by linguistic creatures, but sometimes he applies a wider conclusion: that thinking about thoughts requires language. Indeed, this is how he sums it up in his motto: no
intentional ascent without semantic ascent. However, this conclusion is not warranted by the above argument. As he himself admits, personal-level imagistic thoughts can be the objects of further thoughts (ibid., pp. 163-164). Rather, what cannot be understood outside of language is personal-level inference since any imagistic account of inference (e.g. mental models or mental maps) is derivative from the sentences whose structure they model (ibid.). This should call into question any of the abilities that he rules out for nonlinguistic creatures simply because they seem to require that a thought be taken as the object of a further thought. For example, he thinks that the attribution of propositional attitudes to others is not within the purview of nonlinguistic thought since it would require taking a thought (i.e. the content of the belief) as the object of another thought (i.e. that the person believes/desires the content) (ibid., p. 172). But, again, what the above argument establishes is that language is required for inferences, not that language is required to think about thoughts.

The distinctive role of language, then, is that it enables one to carry out personal-level inferences and hence enables a whole range of operations involving the conscious evaluation of inferential relationships between thoughts. Language is uniquely able to confer this ability because of its structure—the only structure for which we have an idea of what inference is (ibid., pp. 161-164). Thus, for Bermúdez the limitations of nonlinguistic thought have to do with the ability to characterize the inferential relationships between thoughts rather than with the capacity for conceptual thought. Recall that we must, according to Bermúdez, attribute concepts (modes of presentation) in specifying how the nonlinguistic creature represents (ibid., pp. 89, 111, 143, 144).
Bermúdez’s account of thought can be characterized by noting two different senses in which he uses the concept of structure (ibid., p. 111; c.f. 1995, p. 351). On the one hand, there is the question of whether the content of thought is structured and on the other hand there is the question of whether the personal-level vehicles of thought are structured. The minimalist construal of thought answers “no” to both of these questions since it attempts to explain all nonlinguistic thinking without having to posit either structured content (i.e. no subject-predicate judgments) or structured vehicles (i.e. images are the only personal level vehicles available to nonlinguistic creatures). Linguistic thought answers “yes” to both questions since its content is propositionally characterized (i.e. subject predicate sentences) and its vehicles (i.e. linguistic entities) are structured so as to permit inferences. However, Bermúdez argues that there is also a middle ground: we can (and must) attribute structured contents to nonlinguistic creatures and yet the personal-level vehicles of nonlinguistic thought (i.e. images, maps etc.) are not structured so as to enable inferences to be made. Bermúdez’s position can be characterized as follows:

<table>
<thead>
<tr>
<th>(Personal-level) vehicles: logically structured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Content: Structured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Nonlinguistic: Level 2 rationality</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimalism</td>
</tr>
</tbody>
</table>
Minimalism reconsidered

Earlier, I questioned Bermúdez’s characterization of minimalism as excluding any representation beyond the here and now. This may be an accurate characterization of perceptual affordances but, as Bermúdez notes, an account of minimalism is not by any means wedded only to affordances (ibid., pp. 47, 54). If we think about how imagistic reasoning could be employed then it seems there’s no reason to limit minimalist thought to the here and now. I tried to pump this intuition by drawing an analogy between a person searching through a kitchen drawer for something (being guided by visual and kinesthetic images of the problem situation, which is not currently present) and the tool-wielding chimpanzees. Bermúdez thinks that we need to attribute instrumental beliefs to these chimps since he thinks they must be representing contingencies between their actions and the outcome. But we need not understand such behaviors as engaging in the kind of protoinferences Bermúdez recommends if we think about them as using imagistic reasoning instead. Moreover, the explanation according to which we attribute beliefs about contingencies between actions and outcomes would not account for the first time the creature constructed the termite-extracting-wand since at that point there would be no contingency to represent. It is only after the pairing of a successful wand-making-session with a successful termite-extracting-session that there would be a contingency to represent. Thus, unless one were willing to chalk the first wand-making-session up to dumb luck, the behavior must be guided by something other than a representation of contingencies. It seems to me that imagistic reasoning is the obvious candidate.

Another case that Bermúdez discusses in the context of nonlinguistic reasoning is the creature reasoning about predators at the watering hole. As discussed above,
Bermúdez thinks that nonlinguistic creatures are capable of reasoning which carries them beyond the here and now and that they do this by a kind of protologic. In particular, a creature might represent the protocausal contingency: if the gazelle is present at the watering hole then the lion is absent from the watering hole. It can then reason using proto-modus ponens to represent the lion being absent or proto-modus tollens to represent the gazelle being absent. In reasoning such, the creature would go beyond the here and now in that it would not need to have seen the lion, smelled the lion or seen traces of the lion (e.g. lion footprints), etc., in order to represent the presence of the lion. It seems to me, however, that this is a case that is amenable to a minimalist construal.

If the explanadum is the behavior of the creature with regards to its approach of the watering hole, then one thing we would want to explain would be how the creature knows when it’s safe to approach the watering hole. Of course, it knows not to approach it when it sees a lion, but it may approach it even when it is not in the position to see whether there’s a lion there. It is this kind of situation, I take it, that Bermúdez would be interested in explaining since it would require (according to his account) that the creature is doing more than minimalist representation. How might we account for it on minimalist terms? On the one hand, it seems not too unreasonable to assume that some creatures may be designed to see certain situations as dangerous or safe, in which case we could invoke the idea of affordances: the creature who sees a large number of gazelles there sees the situation as affording safety, perhaps. Part of the reason this explanation could work is that the explanadum is not an artificially constructed one. As Bermúdez notes, one can invoke affordances just so long as the behaviors to be explained are not highly artificial ones constructed in the lab (ibid., p. 54). The reason seems to be that
affordances would likely have some sort of adaptive history to be told, but no such
history is in the offing for behaviors such as how a dolphin can learn complex gestural
commands or how a bonobo can use lexigrams to communicate (ibid.).

On the other hand, it may be that the creature learns to see the gazelles’ presence
at the watering hole as indicative of safety and their absence as, perhaps, indicative of
danger or at least licensing caution. However, the end product of this learning might not
be best characterized as a protocausal representation of the contingencies in the
environment. Perhaps what the creature does is represent the situation as being similar to
either a safe situation or a dangerous one, where safe and dangerous situations are stored
in a similarity space (see chapters 1, 2) according to dimensions such as the number of
animals at the watering hole and/or the amount of activity at the watering hole. For
example, situations in which a high number of animals and activity are at the watering
hole would cluster together in a similarity space whereas situations where there were few
or no animals and/or little or no activity would cluster together. The former
would be perceived as safe inasmuch as no instances are associated with or contained any element
of danger (i.e. a lion), whereas the latter would be associated with danger inasmuch as
some of those situations are associated with the presence of a lion. It may be that even a
weak association with danger would be strong enough to prevent the animal from
approaching the watering hole. Such a creature would still be representing contingencies
but these representations would not invoke concepts as Bermúdez’s protocausal
conditionals would. Rather, the representations would be holistic representations of the
overall similarity of the situation in terms of dimensions to which the creature is
sensitive. Thus, the situation would either look safe or look dangerous depending on the
number of animals present. If there were not many animals present, the creature would
exercise much more caution in approaching the watering hole.

_Bermúdez’s account in perspective_

If an account of nonlinguistic thought which does not invoke concepts is possible, then it
may be preferable inasmuch as it would avoid the problem of abstraction, which I have
argued has no straightforward or even promising solution. Bermúdez’s account of
nonlinguistic thought does invoke concepts since we need to attribute them in specifying
the contents (modes of presentation) of their thoughts. Thus, Bermúdez’s account
requires an account of how conceptual representations such as PREDATOR and
ABSENT are created and thus faces the problem of abstraction.

Linguistic and nonlinguistic thought seem to be largely continuous on
Bermúdez’s account: the differences seem to be mostly differences of degree. Bermúdez
has argued for a nonlinguistic analogue of reasoning and inference, the only difference
between the two being that nonlinguistic reasoning and inference cannot operate on
“whole thoughts.” So it is not that nonlinguistic creatures do not reason, it is just that
their reasoning is limited in certain ways and language enables one to overcome these
limitations. One way to understand these limitations is in terms of detachment from the
here-and-now which, as Bermúdez notes, is a matter of degree (ibid., p. 170). Minimalist
thought, according to Bermúdez, is limited to the here and now and one might think that
here is a place where one might note a difference in kind rather than degree. However, I
have argued that there is no good reason to so restrict minimalist thought if one takes into
account imagistic reasoning. Thus, although it may be the case that the level 2 reasoner
can achieve a greater degree of detachment from the here and now than the minimalist thinker, the minimalist thinker can still achieve some detachment from the here and now by manipulating images. Linguistic thinkers on Bermúdez’s account achieve an even greater degree of detachment from the here-and-now in that they can revise their beliefs in the absence of countervailing evidence. Again, it’s not that nonlinguistic creatures cannot revise their beliefs but, rather, that they can only revise them in the presence of countervailing evidence (e.g. the rat’s instrumental belief about the dependence of food on lever-pressing is modified in light of evidence that food will appear regardless of any lever-pressing).

Of course, the ability to revise one’s beliefs in the absence of countervailing evidence brings with it some abilities that are unique to linguistic creatures (i.e. self-monitoring and self-criticism). These abilities stem from the ability to make personal-level inferences, and personal-level inferences must be carried out in a natural language since we have no idea of how to characterize inference with images (the only other personal-level) vehicle. So language does create a new kind of thought on Bermúdez’s account: personal-level inference. Nevertheless, there’s still a sense in which Bermúdez is what in chapter six I will call a continuity theorist regarding the role of language in cognitive development: language helps us to be better thinkers but there is no essential difference between the thought contents of linguistic and nonlinguistic creatures. Both linguistic and nonlinguistic creatures can think thoughts whose representational content is something like THE LION IS PRESENT AT THE WATERING HOLE. Of course, different creatures need not have the same concepts, and Bermúdez also offers an account of how to figure out which concepts a nonlinguistic creature has (i.e. how the creature is
representing the content).\textsuperscript{70} This means that both linguistic and (some) nonlinguistic creatures and have concepts (even if not the same ones). The difference, then, is not so much that the thought of nonlinguistic and linguistic creatures differs in representational content but, rather, that linguistic creatures can become conscious of certain of their thoughts. This contrasts with the account of the role of language in cognitive development that I will offer in the final chapter. On my account language introduces new and different representational contents which are not within the purview of nonlinguistic creatures. On this account there is a discontinuity between linguistic and nonlinguistic thoughts in terms of their representational contents: linguistic thoughts have one kind of content and nonlinguistic thoughts have another.

\textsuperscript{70} I have not discussed Bermúdez’s account of the semantics of nonlinguistic thought here. His account is a success semantics: the content of the belief is given by the state of affairs whose holding would result in the satisfaction of desires, while the content of a desire is its satisfaction condition, i.e. the state of affairs which leads to the cessation of desiring behavior.
Peter Carruthers: Thinking in language

Peter Carruthers (1996, 1998, 2002) has advanced several related claims defending what he calls the cognitive conception of language. He contrasts the cognitive conception with what he calls a communicative conception of language according to which natural language has the function of facilitating interpersonal communication rather than facilitating individual cognition (1996, p. 1). What all of his claims have in common is this: there are some kinds of thinking which can only be done in natural language. So, for example, in both his 1996 and 1998 he defends the claim that conscious, propositional thinking can only occur in natural language. While there may be other kinds of unconscious thought that do not require language, at least conscious propositional thought does.\(^7\) Carruthers 2002, on the other hand, defends the claim that language is required to integrate information across different modules. So, on the supposition that the human mind is largely modular (in that it consists of special-purpose mechanisms which are isolated from each other in that the functioning of one is impervious to the functioning of another), the function of language is to allow different domains of knowledge (encoded in the different modules) to be integrated. Although both of these views share the notion that there are some kinds of thought that can only occur in language, there is a difference. On the earlier view, there are some thought tokens (namely, conscious, propositional ones) which occur in language, but this allows that some unconscious thoughts of the same type (i.e. same representational content) need not be thought in a natural language. On the later view, in contrast, there are some types of

\(^7\) In fact, Carruthers 1996 argues that if one accepts that conscious propositional thought occurs in natural language, then it’s more parsimonious to accept that all propositional thought does (p. 266). But his fallback position is the former, weaker thesis. I will return to these issues below.
thought (namely, thoughts combining information from more than one module) which can only be thought in a natural language. For example, on the earlier view, although it may be the case that I have an unconscious language of thought sentence with the content THAT DOG HAS THREE LEGS, language would be required for me to be conscious of such a thought. In contrast, on the latter view, whereas I may be able to think the thought THAT STUFF IS YELLOW without language, I could not think the thought THAT YELLOW STUFF SMELLS LIKE SULFUR without language.

In what follows, I will examine each of these arguments in detail and will present several problems with Carruthers’’s arguments as well as a counterargument against his theory. A general criticism that will emerge is that although Carruthers is correct in thinking that language has an important cognitive role to play, his acceptance that conceptual thought can occur outside of language undermines his claim that language plays some unique role in cognition.

Carruthers 1996: The argument from introspection

Carruthers 1996 undertakes an ambitious project: he attempts to defend the thesis that it is naturally necessary that conscious, propositional thought requires natural language because it literally occurs in natural language. By way of defending this claim, he develops a theory of consciousness which is supposed to be independently motivated and to stand on its own. According to his reflexive thinking theory of consciousness, a thought is conscious only if both the content of the thought and the fact of its occurrence are available for further occurrent thoughts (1996, p. 194). So, for example, one and the same conscious perceptual state could be used in two different ways: as the thought that,
e.g., THE GLASS IS NEARLY EMPTY as well as the thought that THAT EXPERIENCE MAY BE ILLUSORY (similarly, a photo of the Mona Lisa can be used as a representation of what she looked like as well as a representation the painting itself) (ibid., p. 205). Carruthers’s theory thus requires that not only can one, e.g., think about what one is thinking or perceiving, but one can also think about the fact that one is thinking or perceiving something (ibid., p. 196). (Carruthers defends a higher-order theory of consciousness on pp. 157-163, but I will not recount that argument here.)

His reflexive thinking theory of consciousness is then applied to the case of humans where he argues that language is required for the mechanism of reflexive thought. His main target is Fodor, who holds that all thinking occurs in Mentalese; thus, if Carruthers is to defend his thesis he must eliminate Fodor’s alternative. 72 Carruthers accepts that all propositional attitudes are to be understood as relations to sentences; his question is whether those sentences are Mentalese or natural language sentences (ibid., p. 40). His main line of argument that at least some are natural language sentences is his argument from introspection, and it is this argument that his theory of consciousness is supposed to support. It certainly seems, so the argument goes, that most of our conscious propositional thoughts are entertained by means of images of natural language sentences.

Introspection informs us, in fact, that many of our thoughts are expressed entirely in natural language—for example, my thoughts as I write this, and, no doubt, yours as you read it. When I sit and draft a letter to someone in my head, for example, what figures in my consciousness is a sequence of English sentences in auditory (and perhaps kinesthetic) imagination, rather as if I were dictating that letter aloud. Or, if this is considered too easy for me (since the task is itself a linguistic one), take as an example any case where I sit and try to think through some abstract problem—comparing competing explanations of autism, say. In

72 Since Carruthers is working within the assumption of a realism of mental states (1996, pp. 25-29), he doesn’t propose to take seriously cognitive conceptions of language from anti-realists like Dennett in arguing against Fodor (e.g. ibid., p. 190).
such cases I find that my thoughts will consist almost entirely of inner dialogue (Carruthers 1996, p. 50).

However, the proponent of Mentalese has a ready reply to such phenomenological data: natural language may give us access to our thoughts but may nevertheless not be the medium in which we think. That is, Fodor can accept that images of natural language sentences occur when we think conscious thoughts, he need only deny that these natural language sentences occupy the causal role characteristic of thoughts. But this, so Carruthers claims, flies in the face of “the common sense construal of the introspective datum” since we think that “the imagined sentences which we are aware of occupy the causal roles in our cognition distinctive of thinking” (ibid., p. 61). Of course, there is nothing sacrosanct about common sense, but if it to be overridden then there will have to be some very convincing reasons for thinking that we think in Mentalese rather than natural language. Carruthers spends a great deal of time both criticizing Fodor’s thought based semantics (i.e., providing an account of the semantics of natural language in terms of an independent notion of thought), including his causal co-variance-cum-asymmetric dependence account of representation, and arguing for the plausibility of an alternative language-based semantics (his favored alternative being functional role semantics). The point of all this is to weaken Fodor’s theory so as to strengthen the argument from introspection, since (Carruthers thinks) if Fodor’s theory conflicts with common sense and that theory is not anything close to mandatory or even very plausible, then ceteris paribus the introspective/common sense datum should be allowed to stand (ibid., p. 134). Indeed, Carruthers thinks that all he needs is a draw with Fodor in order for his theory to win, since it is consistent with common sense where Fodor’s is not (ibid., p. 235). I won’t survey Carruthers’s arguments against Fodor, nor will I spend much time
elucidating his reflexive thinking theory of consciousness. Rather, I will take a critical
look at the argument from introspection (on which Carruthers’s case hinges) as well as
the modal claims about the relationship between language and conscious thought he
develops based on it.

Carruthers’s argument that conscious, propositional thought is constituted by
natural language depends on eliminating various possibilities about the significance of
inner speech which are consistent with the idea that all thinking occurs in Mentalese. He
thinks the alternatives fail because they are unable to accommodate the various data as
well as Carruthers’s own reflexive thinking theory which holds that in the case of humans
natural language is required for conscious thought.

The first hypothesis is that language helps to fix thoughts in memory by
associating the Mentalese concept with a natural language word. Perhaps words, as
independent objects in their own right, are associated with the Mentalese equivalent,
making the latter easier to remember (ibid., p. 237). This is not implausible given that it
is clear that associations are beneficial for memory, e.g. associating a word with
something else rather than just repeating the word over and over to oneself better aides
one’s memory of the word. For example, in order to remember a number of different
objects, one might imagine those objects in different rooms of one’s house (ibid.).
Perhaps natural language words have the same effect on helping us to remember
Mentalese words when the need arises. The problem with this idea, however, is that it
does not account for the extent of our inner verbalization. On this account, one might
predict that inner verbalization would only occur only in complex reasoning tasks that are
of some importance (ibid.). However, we find inner verbalization just as often in
unimportant thoughts like day-dreams as we do in complex reasoning tasks. The thesis that we do all our conscious thinking in language, on the other hand, would accommodate this datum.

Furthermore, Carruthers claims that the memory-aid proposal, in supposing that all thinking really occurs in Mentalese, is in danger of implying that we never engage in conscious thinking. The reason is that on his view, conscious thought must be direct and noninferential, but the memory view, in claiming that inner verbalization is not constitutive of the thought itself (since only Mentalese is), is committed to saying that we don’t have non-inferential access to the real (Mentalese) thoughts.

According to the memory-aide proposal, on the other hand, the inner verbalizations to which we have non-inferential access are not our thoughts themselves, but rather events which are caused by those thoughts in order to improve memory. Our access to our own thoughts is therefore not immediate, but inferential—to know what we have thought, we have to decode back again to reach the appropriate sentence of Mentalese which is (according to the proposal) the thought itself. So the thoughts which get verbalized in inner speech are not conscious ones, even though they are thoughts which we may reliably know ourselves to have, much as I may reliably know of the thoughts of other people on hearing them speak (ibid., p. 238, c.f. pp. 243-244).

It seems to me, however, that it’s not at all clear that Carruthers can level this claim against the proponent of Mentalese since the “inferences” that are taking place are sub-personal, not personal. There may be a number of different sub-personal processes that can be characterized inferentially (say in low level vision) but we don’t think that this counts against our perception being conscious. 73 The fact that there may be edge detectors that make certain inferences about the presence of edges based on the presence

73 Note, moreover, that Carruthers talks about inner verbalization being caused by Mentalese as well as it being an inferential process. But does a causal process really undermine introspective immediacy? I will return to this point below when considering Carruthers’s 1998 argument that we think in language.
of certain kinds of retinal information does not seem to undermine the fact I am seeing, e.g. a table. Moreover, it is not clear that Carruthers has characterized the memory-aid position correctly in criticizing it: why need we “decode” from Mentalese to natural language words? It seems, rather, that the proposal on the table would be that words become associated with Mentalese terms and that this association aids memory. But it does not seem that we have to decode Mentalese into natural language words for this to work. One does not need to “decode” the rooms of one’s house into the objects one is trying to remember. I am not sure that it even makes sense for the Mentalese proponent to accept the talk of decoding since according to him, the English words in inner verbalization would not carry any content such that they could be decoded. Rather, they are supposed to function (on the proposal at hand) as a mere association. Thus, it seems that Carruthers’s second criticism of the memory-aid proposal is less than convincing.

Another proposal that Carruthers needs to argue against in order to establish his conclusion that all conscious, propositional thinking occurs in language is that there can be purely propositional thinking—thinking which does not involve inner verbalization. Some people report having purely propositional thoughts, that is, conscious thoughts which are not accompanied by any kind of inner verbalization. Carruthers’s tactic here is to just deny that such thinking exists. In support of this he cites some of the well-known social psychology research on confabulation (e.g. Nisbett and Ross, 1980). For example, when presented with a number of identical items, people will always choose items from the same side but will justify their choices with some rationale (e.g. that the one they picked had a superior quality). The best explanation of cases like this and many others like it is that people lack access to the thoughts which are actually determining their
actions and they engage in a swift bit of retrospective self-interpretation using the
generalizations of folk-psychology (e.g. people generally choose the items they think are
the best). Since such thoughts involve interpretation, then, they would not count as
conscious on Carruthers’s account of consciousness since they are discovered
inferentially, and so are no threat to his thesis that conscious, propositional thoughts are
constituted by images of natural language sentences.

Carruthers offers several cases from Russell Hurlbut in support of the idea that so-
called purely propositional thought is actually a swift bit of self-interpretation. Hurlbut
would outfit people with a device that beeped randomly throughout the day, at which
point the subjects were supposed to report what they were thinking.

In another example, the subject was looking at a box of breakfast cereal on the
shelf of a supermarket at the time of the beep. She reported that she was
wondering—wordlessly—whether to buy the box; and that she was thinking—
again wordlessly—that she did not normally eat breakfast, and that the cereal
might therefore be wasted; and that she was also considering the expense involved
(see Hurlbut, 1993, p. 94). Again, it seems reasonable that these thoughts might
have been ascribed inferentially, as a result of self-interpretation, rather than
occurring in such a way as to render them properly conscious. For these are just
the thoughts which an observer might naturally attribute to her, who knew what
she knew: namely, that she was attending to the price-label on a cereal packet;
that she did not normally eat breakfast; and that she was generally careful in

Of course, even granting all of this, it still does not follow that if we sometimes
confabulate then we always do (thus leaving an out of the defender of purely
propositional thinking). Carruthers’s response to this is that if the swift self-
interpretation explanation is likely in some cases of purported purely propositional
thinking, then a quick parsimony argument/inference to the best explanation establishes
that it likely happens in all cases (ibid., p. 243).
The final alternative proposal that Carruthers considers is that language gives us access to our thoughts even though those thoughts occur in Mentalese. The problem with this proposal, according to Carruthers, is that, like the memory proposal, it implies that we never engage in conscious, propositional thinking. Again, he claims this because according to his reflexive thinking theory of consciousness, a thought is conscious only if one has noninferential access to it. But if our thinking really occurs in Mentalese and must be somehow encoded into images of natural language sentences in order for us to be conscious, then we do not have noninferential access to our thoughts, thus, according to RT theory, we aren’t conscious of them (ibid., pp. 238, 239, 244). Above I questioned whether this was a fair criticism to level against Fodor since its not clear (at least not without argument) that subpersonal causal processes which can be characterized as inferential (which is surely how such “encoding” would be understood) would undermine the status of a mental state as being conscious. There seems to be a significant difference between engaging in inferences about what someone is thinking based on their behavior, on the one hand, and having certain of our nonconscious thoughts cause conscious thoughts, on the other. I will return to this criticism again below.

Carruthers’s argument from introspection thus depends on ruling out alternative explanations of the role of inner verbalization which would support the conclusion that our conscious thoughts are not constituted by natural language sentences. In a moment I will criticize the basic premise on which it rests: that introspection actually gives us the data that Carruthers claims it does. Before doing that, however, I want to examine the modal theses that Carruthers attempts to defend. The argument from introspection, if allowed to stand, shows that our conscious, propositional thoughts are constituted by
images (auditory or other) of natural language sentences. But this leaves open whether it is necessary that they do. Why, for example, couldn’t we entertain such thoughts in some other medium, for example, images? Carruthers thinks that it is naturally necessary that conscious, propositional thinking occurs in language in the sense that given the architecture of human cognition as well as causal laws, it couldn’t be otherwise that our conscious, propositional thoughts occur in language (ibid., p. 252).

Carruthers argues for the natural necessity of language for conscious thought by an elimination of alternatives (images, Mentalese). He first considers the hypothesis that the although some conscious, propositional thinking could be done in language, nonetheless language is not required since the same thoughts can also be thought imagistically. For example, could I also not think the thought “the cat is on the mat” in images which were themselves available to further conscious thinking (as his reflexive thinking theory of consciousness requires)? Carruthers’s answer for this depends on a kind of functional role semantics according to which mental contents are individuated by their causal/functional/inferential interactions with other mental states in their causing of behavior (ibid., pp. 111-118, 123-127). He notes that images and sentences differ in that images cannot capture the same meaning as a sentence because it will always have “excess content” (ibid., p. 255). For example, an image of a cat on the mat will provide more than the sentence does such as whether the cat is sitting or standing on the mat (ibid., p. 38). But then the image would play a different role in cognition, in which case an image could never play the same role as a sentence. Thus, a conscious image could never do the duty of a conscious sentence, in which case Carruthers’s thesis stands. In
short, functional role semantics plus different functional roles for images and sentences equals unique linguistic contents.

Mentalese is not up to the task either but for a different reason than images. Images are conscious but have different contents than sentences; Mentalese sentences would have the same contents as sentences but their problem is that we do not have conscious access to them (ibid., p. 256). It might also be thought that Mentalese itself has some sort of semantic ascent mechanism, in which case certain thoughts of Mentalese could be conscious. The problem with this proposal (as was noted above) is that much of the evidence we have for purely propositional thinking (i.e. those where neither sentences or images are used) suggests that we actually confabulate that such episodes: we posit them after the fact in a swift self-interpretive act (ibid., p. 257). Moreover, even if we were to suppose there were such a Mentalese semantic ascent mechanism, it would be an extravagant hypothesis if one had already accepted that in some cases language is used for reflexive thinking (ibid.).

Thus, it seems that neither images nor Mentalese can play the role of natural language sentences in conscious, propositional thought. In that case the weak thesis of the natural necessity of language for thought stands. Carruthers also suggests a strong thesis of natural necessity is plausible given the acceptance of the weak thesis. The strong thesis states that not only do some thought tokens occur in language (i.e. conscious, propositional ones) but, further, that some thought types do. In particular, the strong thesis of natural necessity claims that not only does conscious propositional thinking occur in language (allowing that these same thoughts can be entertained nonconsciously without language) but that some kinds of propositional thinking tout
court occurs in language. Carruthers’s suggestion is that the types of propositional thought that require language are those whose concepts depend on the acquisition of language (ibid., p. 271). He leaves open that there may be many concepts which do not depend on public language for their acquisition and that may, therefore, be outside the scope of the strong thesis of natural necessity. He includes in this category concepts of so-called folk-physics and folk-biology, but cautions that the abilities characterized in the developmental evidence for these categories may not be best characterized as thought (i.e. as conceptual thought). I am in agreement with him here: this is, in fact, what I argued about the child’s object concept (contra Xu in chapter 3) as well as with regards to prelinguistic sortal concepts (contra Mandler in chapter 2).

The main problem with Carruthers’s argument from introspection is that the introspective data on which he builds his case are far from obviously true. The original claim was just that it is characteristic of our conscious thought that we are aware of it as images (auditory or kinesthetic) of natural language sentences. However, given that Fodor and proponents of Mentalese can acknowledge this much, Carruthers goes on to claim that the common sense interpretation of such inner speech is that “the imaged sentences which we are aware of occupy the causal roles in our cognition distinctive of thinking” (ibid., p. 61). Carruthers continues:

Fodor must claim that both I, and common sense, are mistaken. What I am aware of here is not the occurrent intention itself (which is, he will say, expressed in Mentalese), but some mere concomitant of it. My dispute with Fodor can thus concern the causal roles of imaged natural-language sentences, rather than their existence. The dispute can be about whether natural-language sentences occupy the causal roles within our cognition of thinkings, decidings, and wonderings whether; not about whether such sentences figure in cognition at all. My claim is that it is the imaged sentence itself—‘I shall get it by climbing on the box’—which occupies the causal role of the thought (the thought, namely, that I will get it by climbing on the box). Fodor’s claim must rather be that the thought itself,
expressed in Mentalese, somehow causes the imaging of the natural-language sentence (as well as causing me to climb on the box) (ibid.).

But is it really common sense that our inner verbalizations have causal power? I don’t think this is a matter that common sense takes a stance on. It may be common sense that our thoughts cause our actions, but that is perfectly consistent with everything Fodor says on the matter. Indeed, Fodor is a champion of the reality and causal efficacy of mental states. Whether or not our conscious thoughts have causal powers via natural language sentences or sentences in Mentalese seems, on the other hand, to be an issue far removed from common sense. In fact, Carruthers’s claim is equivalent (or at least implies) that common sense says that we would never have many of our conscious thoughts if we did not have language. But that is far from a common sense platitude! But if common sense takes no stand on this issue, then the main line of argument for the necessity of language for conscious, propositional thought fails. The reason is that Carruthers’s argument is essentially that his theory is a viable one and accommodates the common sense construal of the introspective data, while Fodor’s is a viable theory which does not accommodate it (cf. ibid., pp. 134, 235). But this assumes that common sense takes a stand on the issue which is in dispute between Carruthers and Fodor (whether conscious thought is causally efficacious in terms of natural language or Mentalese), and I don’t think that it is at all plausible that common sense takes a stance on this issue. I will later suggest that Carruthers’s mistake is to concede that there can be conceptual/propositional thought outside of language at all, since this opens him up to a parsimony argument in favor of Mentalese (or some other kind of conceptual thought) over both Mentalese and natural language.
Carruthers 1998 puts forth the following argument that conscious propositional thinking is constituted by natural language sentences:

1. Conscious thinking requires immediate, non-inferential, non-interpretive, access to our own occurrent thoughts, and that access is distinctively different from our access to the thoughts of other people.
2. Occurrent propositional thoughts either receive articulation in inner speech, or they don’t; and if they do, then inner speech is either constitutive of the thought-tokens in question (the cognitive conception), or not (the communicative conception).
3. If the manipulation of natural language sentences in imagination (in ‘inner speech’) isn’t constitutive of propositional thinking, then our access to those of our thoughts that receive expression in inner speech is interpretive, and similar to the sort of access that we have to the thoughts of other people, when they speak; and hence such thoughts of ours don’t count as conscious (by 1).
4. The form of access that we have to those of our occurrent propositional thoughts that don’t receive expression in inner speech also involves self-interpretation, and hence such thoughts, too, fail to count as conscious (by 1).
5. So if we engage in conscious propositional thinking at all, then natural language sentences must be constitutively involved in such thinking (from 1, 2, 3, and 4).
7. So natural language is constitutively involved in conscious propositional thought (from 5 and 6).

The key premises to defend here are one, three and four. His defense of premises one and four draw on his 1996. I have mentioned above, and will not take issue with, his theory of consciousness according to which for a thought to be conscious it must be noninferential—that is, not inferred on the basis of other beliefs (premise 1). And his argument in support of premise four have also been canvassed earlier—namely that purely propositional thoughts (i.e. those not accompanied by inner speech) are really just confabulations based on swift self-interpretations. The premise I want to look at more closely is premise three.
Carruthers argues against the following kind of picture:

First a thought is entertained in a manner that does not constitutively involve natural language (carried by a sentence of Mentalese, as it may be); and then that thought is encoded into a natural-language sentence with the same (or sufficiently similar) content, to be displayed in auditory or motor imagination. By virtue of the conscious status of the latter, we thereby gain access to the underlying thought (2005, p. 126).

The problem with this picture, Carruthers thinks, is that it requires some process of interpretation in going from the (nonconscious) Mentalese sentence to the (conscious) natural language sentence.

In order for me to know which thought it is that I have just entertained, when the sentence ‘I just have time to get to the bank’, figures in auditory imagination, that sentence needs to be interpreted, relying on cues provided by context. These cues can then presumably be both cognitive and non-cognitive in nature. What enables me to disambiguate the sentence may be other recent thoughts, or current goals, of mine; or it may be background knowledge of the circumstances, such as that there is no river near by (ibid., p. 127).

But if interpretation is required, then the thoughts are not known immediately which, by premise one, disqualifies them as conscious. I think the clear place to push here is on the idea that the process from nonconscious Mentalese thoughts to conscious natural language sentences must be a matter of interpretation. Why not say that the Mentalese causes an already disambiguated natural language sentence? Carruthers himself raises this as an objection: “Why shouldn’t the [Mentalese] sentence just drag its own interpretation with it, as it were—acquiring the content, not through any sort of process of inference, but simply by virtue of its causal connection with the underlying thought that it serves to express (ibid., p. 128)?”

Carruthers gives two responses to this objection. First, it seems to be the case that visual imagery, in general, requires interpretation and if this is so then we should view inner speech similarly. According to Carruthers, the best theories of visual imagery see a
visual image being input into the visual system which subsequently interprets it in the normal way (ibid.). Thus, since inner speech is a kind of imagery, it is likely that inner speech likewise produces an input to the language system which then interprets that input (ibid.). Second, since the systems that produce and consume must be distinct, it is unlikely that the content of inner speech can be assigned without some kind of interpretation that relies on the thinker’s current beliefs (ibid., p. 129). Carruthers thinks that the latter response is the more powerful response. However, it doesn’t seem to me very powerful at all. The idea (I think) is supposed to be that it is absurd to think that a mere causal process can create a contentful state. Carruthers notes:

The idea of a sentence ‘dragging its own interpretation with it’ is surely incoherent, indeed. If the mere fact of having been caused by a certain thought were sufficient to confer an interpretation on it, from the perspective of the consumer system, then one might just as well say that the mere fact that my spoken utterance is caused by a particular thought of mine is sufficient for you to interpret it. But that would be absurd (ibid.).

However, this misconstrues the position being argued against. Carruthers’s criticism already assumes what the position he is arguing against denies, namely, that inner speech has its own stand alone content. But this is precisely what the Mentalese proponent would deny: only Mentalese are real thoughts, inner speech is merely a way to access the content of that thought. The illicit assumption in Carruthers’s argument here is that there are two distinct, contentful items: Mentalese and inner speech. This is where the analogy between inter- and intrapersonal speech goes awry: of course speech doesn’t wear its interpretation on its sleeve, but this is precisely the disanalogy with Mentalese on the current proposal, since the Mentalese thought is suppose to cause its conscious interpretation in inner speech. Although this would be an absurd model for interpersonal speech, it does not seem to be absurd for an intrapersonal cognitive process. Thus, the
proponent of Mentalese should just deny that going from the Mentalese sentence to inner speech requires interpretation. But in that case, premise three is false (or at least not obviously true without further argument) and so the argument is unsound.

**Carruthers 2002: The argument from cross-modal thought**

Carruthers has put forward an argument in support of the idea that there are some types of thought that can only occur in language. Although this argument does not depend on considerations about conscious and nonconscious thought, it does require a commitment to the modularity thesis of human cognition. The modularity thesis supposes that many of the cognitive abilities humans have are the result of distinct, functionally specified, evolutionarily-driven brain structures whose processing is “informationally encapsulated” (i.e. the information available to one module is unavailable to another module and vice versa). A hackneyed example of such informational encapsulation is optical illusions, where even when we know that, e.g., the two lines are the same size, the one still looks longer than the other (as in the Mueller-Lyer illusion). It seems here that our perceptual processes are isolated from our beliefs so that the former cannot utilize the deliverances of the latter.

Carruthers supposes that our minds contain a number of different modules (e.g. naïve physics, naïve psychology or mind-reading, folk-biology, intuitive number system, a system for keeping track of social contracts) (Carruthers 2002, p. 663). Each of these modules can generate (conceptual) thoughts about their respective domains. For example, a geometric module might generate the thought THE TOY IS IN THE CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE
RIGHT, whereas an object-property module might generate the thought THE TOY IS BY THE BLUE WALL (ibid., p. 669). The problem is that the information cannot be shared between the modules, and this, Carruthers supposes, is the function of language: to share information across modules so that one can think propositional thoughts like THE TOY IS IN THE CORNER WITH A LONG WALL ON THE LEFT AND A SHORT BLUE WALL ON THE RIGHT. As Carruthers notes, this hypothesis is briefly suggested by Dennett (1991, pp. 196, 301, 315) and Spelke (2003). This thesis differs from Carruthers’s earlier one (viz. that conscious propositional thought occurs in language) since it claims that there is a certain type of thought that can only occur in language, regardless of whether that thought is conscious or nonconscious.

In support of this hypothesis Carruthers cites several lines of empirical evidence. First, there is indirect evidence from cognitive archeology which suggests that cross-modular thinking came about at the same time as language (ibid., pp. 665, 667). For example, although Homo erectus and early Homo sapiens could make complex stone tools, they did not make specialized tools for hunting different kinds of game, nor did they make tools out of other materials such as antlers and bones—thus they did not integrate knowledge from different domains. Second, and more significantly, Carruthers cites several lines of direct empirical research which suggest that cross-modular thinking occurs in natural language. First, he cites work by Hermer-Vazquez et al. 1999 which draws on prior work about the potential existence of geometric module in rats in which rats, while capable of noticing non-geometric properties (e.g. the colors or patterns of walls or a strong scent) were nonetheless only able to use a geometric information in finding their way through a maze to the food. Hermer and Spelke 1994 found a similar
kind of thing with prelinguistic children: they were incapable of integrating geometric and non-geometric cues in searching for an object, and the only significant predictor of success was spontaneous language usage (cited in Carruthers 2002, p. 667). Hermer-Vazquez et al. 1999 set out to see if the same kind of result could be found with adults. Since adults have language they used a shadowing task: adults had to solve a task which involved integrating geometric and non-geometric properties while repeating back speech that was being played to them through headphones (cited in ibid.). There was also a control condition in which the subjects had to play a rhythm rather than repeat speech. The results showed that the speech repeating group had a much more difficult time solving the problem than the rhythm repeating group (ibid.).

But how might language enable encapsulated modules to share information? Carruthers suggests that cross modular thinking is made possible by linguistic representation which is stripped of all its imagistic-phonological features, similar to Chomsky’s level of representation he calls “logical form” (ibid., p. 666). By hypothesis, this language faculty has access to the outputs of the various modules and so can combine information across domains (ibid.). But what of speech comprehension? As Carruthers notes, there is a prima facie problem in understanding both how the modular information could be combined as well as how the contents of speech could be split up so as to be understood by the modules again (ibid., p. 670). It is in answering this question that Carruthers encounters a problem. He suggests that “mental models” may be needed in speech comprehension:

When listening to speech, what people do is construct a mental model of the situation being described, which they can then use to underpin further inferences. The reason why this may work is that mental models, being perception-like, are already of the right form to be taken as input by the suite of conceptual modules.
For of course those modules would originally have been built to handle perceptual inputs, prior to the evolution of language (ibid.).

This claim, I propose, creates problems for Carruthers, for if mental models are needed anyway for speech comprehension, then logical form sentences of language seem superfluous. It seems that mental models could perform the integration necessary for domain-general thoughts. This is especially pressing given that Carruthers admits that there is conceptual thought that takes place outside of the logical form sentences of the natural language system, namely within the modules (ibid., pp. 661, 661, 666-667). The picture seems to be something like this:

Module 1 \[\rightarrow\] LF \[\rightarrow\] Mental models \[\rightarrow\] Module 1
Module 2 \[\rightarrow\] LF \[\rightarrow\] Mental models \[\rightarrow\] Module 2

My question is: why is logical form needed at all? It seems it isn't, for if the modules can interpret (in speech comprehension) a mental model into their proprietary, domain-specific concepts, then there seems no reason to deny that the modules could communicate with each other using mental modules, without having to utilize the logical form of natural language sentences. Thus, non-domain specific thinking would be possible outside of natural language.

One response to this objection might be that mental models do not carry propositional content, rather, the logical form sentences do. The mental models are needed only to translate the (propositional) logical form sentences back into the proprietary language of the modules. However, my point is that the modules in this instance can communicate with each other since they share the mental model as an intermediary. It is another question whether mental models can carry propositional content. It seems to me that to the extent that Carruthers is willing to countenance
conceptual/propositional thought outside of language (in the modules), this sets him up for potential responses from, say, defenders of Mentalese who would want to claim that domain general thinking is carried out not by logical form sentences of a natural language but, rather, by Mentalese sentences. Another response Carruthers might make is to invoke the argument from introspection that much of our thinking is, in fact, carried out in natural language sentences. However, I have argued above that the argument from introspection is unsuccessful.

In addition, Philip Robbins raises a key objection to Carruthers’s account: it seems that Carruthers already has domain integration within the modules, for how else could e.g., the geometric module construct the thought THE TOY IS IN THE CORNER WITH A LONG WALL ON THE LEFT AND A SHORT WALL ON THE RIGHT (2002, p. 697)? This thought, after all, includes non-geometric concepts such as TOY and WALL (ibid.). Moreover, the mindreading module would have to have access to other nonpsychological domains in order to attribute the thought contents and Carruthers accepts that language is not necessary for this kind of cognition (ibid.). However, if we move to a stronger notion of domain-integration in order to provide a role for language, such as being able to make inferences between modules, then it becomes increasingly implausible that the language faculty should have a general-purpose inference engine (esp. on the assumption of modularity, which Carruthers accepts) (ibid.).

**An argument contra Carruthers from parsimony**

If Carruthers’s argument from introspection and argument from the noninferential nature of conscious thinking fail, this leaves him open to a kind of parsimony argument against
the idea that language is constitutively involved in thinking. I have argued above that there is no common sense construal of inner speech as causally efficacious, but if so then the introspective data about inner speech is just as consistent with Fodor’s theory as it is with Carruthers’s. This creates problems for Carruthers in light of his acceptance of conceptual thinking which is independent of language. The reason is that if he is willing to accept nonlinguistic (nonconscious) conceptual thinking, then it would be simpler to suppose that all conceptual thinking takes place in a nonlinguistic format (e.g. in Mentalese or connectionist-style representation). That Carruthers does think there is nonlinguistic conceptual representation is clear for a number of reasons. First, the knowledge stored in modules is conceptual (2002, pp. 661, 662, 666-667). Second, he admits that animals and infants have conceptual thoughts (albeit non-conscious) (1996, pp. 220, 221). Third, he admits that although conscious propositional thinking requires language, it may very well be the case that non-conscious propositional thinking occurs in Mentalese (1996, p. 66). In light of the failure of the argument from introspection, then, I suggest that Carruthers’s acceptance of nonlinguistic conceptual representation (whether Mentalese or connectionist-style) leaves him open to the charge that it would be simpler to suppose that all conscious propositional thinking is independent of language and not constituted by it.

Carruthers’s option here would be to either deny that Mentalese is a viable account of conceptual representation at all (not merely that it’s not a good account of conscious propositional thought) or admit that there are two separate systems of representation: Mentalese and natural language-based sentences. (Of course, he could also accept that all thought occurs in Mentalese or some other medium independent of
natural language, but I assume he would not want to do that.) There may, of course, be
two different systems of representation but if there is no reason to posit two separate
systems of representation, then why do so? It seems to me that the best option here
would be to argue for the stronger thesis that Mentalese is not a plausible account of
conceptual thought *tout court* (not merely that it isn’t a good account of conscious
conceptual thought). If Carruthers would do this then he would not be susceptible to this
parsimony argument since he would have ruled out the alternate conceptual
representation system which is competing against his natural language representation. To
be fair, Carruthers does argue against Fodor’s causal co-variance theory of representation
as well as against the Gricean theory of linguistic communication which Fodor’s theory
implies (1996, chapter 5). But he is still willing to allow that nonconscious thinking is
done in Mentalese and this is what creates the current problem, given the failure of his
argument from introspection.

It seems, then, that the acceptance of conceptual representation which is
independent of language ultimately creates problems if one is attempting to defend a
strong cognitive conception of language (as Carruthers is). The reason is that if you
already accept a nonlinguistic representational system which is independent of language
then there seems to be a serious hurdle to get over (namely, parsimony) in order to defend
the thesis that language is somehow constitutive of a certain kind of thought. I suggest
that the best way to defend the cognitive conception of language is to argue that there is a
type of representation that is simply unavailable to nonlinguistic creatures. This is what
Carruthers 2002 attempts to do but, as we have seen, the type of thinking which he claims
is unique to linguistic representation (namely, cross-modal, domain-general thinking)
does not seem to require language at all. In the last chapter I will suggest a better way of defending the cognitive conception of language: linguistic representation makes possible a new kind of representation—a kind of representation not within the purview of nonlinguistic creatures.

I think that what Carruthers’s account suggests is that we are probably not going to find a persuasive reason to attribute a cognitive function to language as long as we focus on intrapersonal uses of language (e.g. making thought available to consciousness, making cross-modular thought possible). The reason is that language could always be the clothing in which our thought appears to us instead of actually constitutive of our thinking. This is the problem that was raised over and over in my criticism of Carruthers. It is only when we begin considering the interpersonal effects of language that we can account for the cognitive functions of language. In the final chapter I will develop an account along these lines.
In chapters one and two I presented a distinction between conceptual and nonconceptual thought. An example of a basic conceptual thought is thinking of a particular as belonging to a kind, whereas to think a nonconceptual thought is to be able to make certain kinds of discriminations without thereby being able to think of a particular as belonging to a kind. Whether or not the various theorists I have discussed would accept my definitions, there must be some kind of simpler (nonconceptual) roots out of which conceptual thought develops, or is learned. Unless one accepts that conceptual thought is unlearned, it seems that there are two options regarding the relationship between conceptual thought and its nonconceptual precursor: either conceptual thought is continuous with nonconceptual thought or it is discontinuous with it.

The continuity thesis would claim that conceptual thought can be built out of the resources of nonconceptual thought. This is not to deny that there may be some important bootstrapping devices, including language, that are involved in the development of conceptual thought. Rather, it is just to claim that the meaning or representational capacity of conceptual thought is ultimately constituted from simpler, nonconceptual resources. Both Barsalou and Mandler exemplify the continuity thesis in that they attempt to provide an account of how conceptual thought emerges from and is constituted by representations which are themselves nonconceptual. For example, Mandler’s image schemas are the simplest conceptual representations but are themselves created out of nonconceptual perceptual representations in conjunction with certain innate attentive proclivities and the mechanism of “perceptual meaning analysis.” Barsalou, on
the other hand, thinks that concepts are groupings of (nonconceptual) “perceptual symbols” and that these groupings (what he calls “simulators”) are established by some sort of grouping procedure according to which similar perceptual symbols are stored in the same simulator. In chapters one and two I argued against continuity by presenting what I called the problem of abstraction. In addition, throughout the dissertation I have suggested that a lot of work can be done by nonconceptual thought. In particular, behaviors of prelinguistic children which are sometimes taken as evidence for them having concepts can be given alternate explanations in terms of nonconceptual representations.

The discontinuity thesis, in contrast, would claim that the representational resources of nonconceptual thought are different in kind from the representational resources of conceptual thought and thus cannot ground conceptual thought. One might wonder how the discontinuity thesis can even get off the ground: How can conceptual thought be sui generis and not build on the resources of nonconceptual thought at all? Ex nihilo nihil fit, after all. One way of attempting to defend discontinuity would be by saying that conceptual thought is discontinuous with nonconceptual thought because to acquire conceptual thought is to learn a natural language. Conceptual thought doesn’t come from nowhere, it comes from language (so the claim would go). This is consistent with maintaining that nonconceptual resources may be used in arriving at conceptual thought. The discontinuity claim refers only to the fact that conceptual representation cannot be built up solely from those of nonconceptual thought. So, although we may use nonconceptual thought in learning a language (for example, bringing our pattern-recognizing capacities to bear on recognizing words as sound patterns), what we have
once we have learned a language cannot be understood or accounted for using only the representational resources of nonconceptual thought.

Although I will not argue for the preceding picture in this dissertation, I do want to suggest a way to vindicate another kind of discontinuity—what I will call discontinuity_L. The discontinuity_L thesis claims that language makes possible new kinds of representation. Thus, prelinguistic and linguistic thought would be discontinuous with respect to their representational capacities. In contrast, the continuity_L thesis would claim that language does not make possible new kinds of representation. Thus, prelinguistic and linguistic thought would be continuous with respect to their representational capacity.

All the of the theorists I have engaged in chapters 3-5 want to see language as having an important cognitive role and most of them set out to vindicate some version of discontinuity_L. In this chapter I argue that none is successful and that they all share a common assumption about how language might increase the representational capacity of thought. Perhaps in light of their failure to vindicate discontinuity_L we should question this assumption and look to a different kind of way of vindicating discontinuity_L.

Thus, I have three tasks for this chapter. First, I will examine whether the views of the theorists I have presented in previous chapters fall within the continuity_L or discontinuity_L thesis. My interest in picking the philosophers and psychologists that I have is that these are people who are attempting (rightly, in my view) to provide an important role for language in cognitive development. I will argue that all of these theorists fall within the continuity_L thesis of cognitive development and that as a result they are not able to see language as truly fundamental to cognitive development. Second, I will diagnose why these theorists are unable to ascribe to language this role. In
particular, I will argue that although they see language as important intrapersonally, they
do not take into account the interpersonal role of language, and that it is this interpersonal
role that is key to the real power of language to create new kinds of thought. Finally, I
will offer in support of the discontinuity thesis an example of a way in which language
makes possible a kind of thinking which is discontinuous with prelinguistic thought.

**Continuity or discontinuity? An assessment**

Andy Clark and Daniel Dennett give similar accounts of how language might possibly
play a role in creating more powerful representational capacities. Their theories are
similar with respect to the function that words play: words function as found objects
which become linked to preexisting representations. As objects, words can be taken as
objects of further thought, which opens up the realm for meta-thinking: thinking thoughts
about other thoughts. The crucial question that I raised for both Dennett and Clark was:
what is the relationship between words and the preexisting, nonlinguistic representations?
It is clear enough how manipulating written symbols can be an aid for memory-limited
creatures, and hence can facilitate our thinking. What is not clear on either Dennett’s or
Clark’s view is how words could bring with them new representational powers.
Inasmuch as words merely “stand in for” more complex representations “already laid
down in the system,” it seems that words are not creating any new representational
capacities. Rather, they are merely serving as mnemonic tools for memory-limited
creatures. But since language does not really create any new representational capacities,
both Dennett’s and Clark’s accounts of cognitive development fall squarely within the
continuity thesis.
José Bermúdez’s account of the role of language in cognition is that language makes possible second-order cognitive dynamics such as the ability to reflect on and evaluate the inferential relationships between the premises and conclusion of an argument. So, for example, linguistic creatures can come to accept or reject a belief without having to have currently present evidence for or against it. Instead, we can sometimes reject or accept a belief because, upon reflection, we realize that there either is or is not good evidence for it. Thus, the ability to make personal level inferences is unique to language using creatures, according to Bermúdez. If personal level inferences are indeed unique to linguistic creatures, then Bermúdez’ account would seem to be a kind of discontinuity thesis: language makes possible personal level inferences, which cannot be represented in a nonlinguistic format.

However, there are some reasons to doubt that Bermúdez really should be seen as a discontinuity theorist. First, consider the fact that Bermúdez does not think that linguistic and nonlinguistic creatures differ in the conceptual thoughts they can think. Both linguistic and nonlinguistic creatures can think conceptual thoughts like A GAZELLE IS PRESENT AT THE WATERING HOLE. The difference is just that linguistic creatures can use this thought in a personal level inference whereas nonlinguistic creatures cannot. Burmúdez’s argument against nonlinguistic creatures being able to engage in personal level inference is just that the only other conscious vehicle we know of (images) does not have the right structure to capture inferential relationships. But nonlinguistic creatures can nevertheless engage in inferences, it is just that they are subpersonal.
This leads me to question whether for Bermúdez the function of language for cognition really has anything to do with altering the representational power of thoughts rather than with making certain thoughts conscious (a view that begins to look like Carruthers 1996). There may be many interesting abilities that consciousness brings, but it need not bring new representational abilities. For all Bermúdez has shown, he has not shown that inferences of the very same sort that occur on the personal level using language, might nevertheless still occur at the subpersonal level. This seems a live option given that he admits that both conceptual thought and protoinference take place in the absence of language. But if the only thing that really differentiates linguistic and nonlinguistic creatures is their ability to engage in conscious inferences and not inferences tout court, then Bermúdez would not really be a discontinuityL theorist after all. Rather, he would be a continuityL theorist since the same representational power of inference could be available to a nonlinguistic creature (albeit at a subpersonal level).

That is, for all Bermúdez has shown it need not be the case that inference is unique to linguistic beings, but only that conscious or personal level inference is. But this is consistent with claiming that the representational power of our thoughts is independent of language. Whether or not consciousness somehow depends on language remains an interesting question, but one which I am not addressing.

Peter Carruthers claims that there are certain kinds of thoughts that can only occur in language: conscious, propositional thought and cross-modal thought. In the previous chapter I argued that Carruthers has not established that either kind of thought requires language. The claim that conscious, propositional thought requires language is straightforwardly compatible with the continuityL thesis since Carruthers is willing to
endorse the claim that the very same representational contents are available to nonlinguistic creatures. The claim that cross-modular thought requires language, however, falls within the discontinuity_L thesis. Supposing, for the sake of the argument, that the mind is made up of modules which cannot communicate with each other, language is supposed to play the key role of allowing the modules to talk to each other. (As I noted earlier, Dennett 1991 briefly proposes a similar idea.) If language really did make cross-modular thought possible, then linguistic thought would bring with it new representational powers which are not within the purview of nonlinguistic thought.

However, Carruthers’ account is not able to establish that language is needed for cross-modular thought. In particular, he has not shown why any kind of intermediary is need for modules to communicate with each other since by hypothesis the modules can read the multimodal information that is supposed to be captured in Carruthers’ LF (see chapter 5). Moreover, the fact that he seems to assume in places that modules can represent cross-modal information undermines his own argument that they cannot. The basic problem with Carruthers’ account is that once he has admitted conceptual thought outside of language, the specter of irrelevance haunts all his attempts to provide a cognitive role for language. The defender of mentalese has too many easy rebuttals to Carruthers’ view. In the previous chapter, I suggested that the problem is that Carruthers only considered the intrapersonal functions of language and neglected the interpersonal functions. I will develop this idea further in the next section. Before moving on to that, however, I want to consider whether Fei Xu and Jill de Villiers are continuity_L theorists or discontinuity_L theorists.
Susan Carey and Fei Xu have argued that infants have an object concept which accounts for certain kinds of object tracking skills they have. In chapter three I argued against this interpretation of the data, claiming that there were other nonconceptual interpretations of these tracking skills, that the so-called object concept lacks certain key features of concepts, and that to suppose that infants are using a concept leads to indeterminacy puzzles regarding the application of the concept. Fei Xu has further argued that children have a sortal concept SPELKE-OBJECT before they have kind concepts like DOG and that language plays a key role in the transition from the undifferentiated SPELKE-OBJECT to the various different kinds concepts (DUCK, BALL, etc.). I think it is clear that Xu is a continuity theorist for the following reason:

Words according to Xu function so as to lead children to expect that the same words will apply to objects that share certain properties. Thus the application of the same word to different objects leads the child to look for properties that those objects have in common. The child then forms the kind concept based on the properties that these objects have in common. In Xu’s terms, words are “essence placeholders.” But, according to Xu, although words play a key role in the learning of kind concepts, the kind concept is due to language-independent mental mechanisms. Ultimately, this is some sort of abstractive mechanism whereby the mind groups together objects according to those properties which are supposed to be definitive of the kind. Thus, language acts like an attention grabbing device: Children, confronted with the blooming, buzzing confusion of objects with different properties, attend specially to the properties of objects that are called by the same word and then abstract the concept based on those shared properties. Words are essence placeholders but that essence, once formed, is a concept which owes none of its
representational powers to the word. The word functions as a bootstrapping device or ladder which can be kicked away once the goal (kind concept) is achieved. Xu is both a continuity $C$ and continuity $L$ theorist because on her account the representational powers of full-fledged conceptual thought are built out of the representational resources of the earlier stages. Importantly, language does not introduce a break or discontinuity because language itself, on Xu’s picture, is not contributing to the increased representational power. Rather, its role is the guiding of attention to those bits of the world that the child’s community has some reason to group together. These groupings are then created by language-independent mental mechanisms—the “essence” is formed.

Jill de Villiers’s “linguistic determinism” attempts to provide a key role for language in enabling children to pass the false belief task. According to her account, there is a certain class of syntactic structures that natural languages contain and these structures enable the child to represent false complements, which are key in being able to attribute false beliefs. What, then, is the relationship between the ability to represent false complements and the syntactic class which gives a structure to represent them? De Villiers certainly claims that one cannot pass the false belief tasks unless one has acquired these syntactic structures, but this leaves open how the syntactic structures make representing false complements possible. On the one hand, we can imagine that, much as on Xu’s account, the particular linguistic structures draw our attention to some underlying conceptual distinction, which our minds then represent independently of language. On the other hand, de Villiers’s account might claim that somehow the syntactic structure of natural language sentences is the very medium in which we think. De Villiers has explicitly acknowledged that she intends to leave open whether or not we
actually think “in language.” I am sure this is a good tactical move on her part since it would be quite a project to try to give an account of the sense in which language could be thought of as the medium of thought. Given that de Villiers’s account could be interpreted in either of these two ways, it is difficult to determine whether she is a continuity\(_L\) or discontinuity\(_L\) theorist. She thinks that language is necessary for one to pass standard false belief tasks—hence her “linguistic determinism”—but this is consistent with the continuity\(_L\) thesis since language need not be upholding the ongoing increased representational power. Rather, it might be a ladder that is kicked away. The discontinuity\(_L\) theorist, in contrast cannot kick the ladder of language away since it continues to be a supporting structure of the new representational powers of the thought it creates. A better metaphor for the discontinuity\(_L\) theorist would be that of a foundation rather than that of a ladder.

**Intrapersonal functions of language: A common assumption**

In this section I want to suggest that one of the features common to all of the theorists I have surveyed is that they only look for cognitive effects of language within the individual. That is, in attempting to provide a role for language in cognition and cognitive development, they seem to assume that the only relevant effect that language might have is within the cognitive economy of the individual. I will contrast this intrapersonal account of language with an interpersonal account according to which the real effect of language on thought derives from its interpersonal functions.

Carruthers 1996 is one who exemplifies the intrapersonal language approach. Although he struggles mightily to provide a cognitive role for language, in the end, as I
argued in the last chapter, he is unable to secure any good reason for the claim that
language makes possible new kinds of thought. One of the main problems his account
faces, as noted earlier, is that having admitted the language of thought into his account, it
could always be the case that language is merely the clothing in which the real thinking
(i.e. Mentalese) appears to us. I think that the problem that Carruthers 1996 faces is
really a manifestation of a more general problem: How can a merely intrapersonal use of
language create a more powerful representational system? I will try to make this problem
perspicuous.

If we suppose, with most, if not all of the theorists I have discussed, that there
exists a conceptual representational system that is independent of language, then how can
natural language combine with that system to create new and more powerful
representations? If we suppose that there really is a nonlinguistic, conceptual
representation system, then it seems there is no compelling reason to deny that there are
some kinds of thoughts that are in principle inaccessible to that system. The philosophers
I have looked at attempt to provide such a reason, but ultimately their attempts fail to
show what is unique to language in creating new representational powers. Carruthers and
Bermúdez, for example, in effect each claim that the uniqueness of linguistic thought has
more to do with consciousness rather than representation (or so I argued above).
Carruthers also makes a claim about cross-modal thought, but this attempt fails as well,
as I argued in the previous chapter. Clark and Dennett, on the other hand, attempt to
make claims about the increased representational power of thought that words enable, but
it is not at all clear even on their own construals that words really do create new
representational capacities (as opposed to alleviating memory and processing limitations).
Of course, one could say that there are natural limitations on our memory and processing capacities, and that language helps to alleviate processing and memory strains and in this way has an effect on what we are able to represent. But one could admit this much and yet maintain that language brings no new representational properties to thought (as Fodor 1975 does). I think Fodor is close to correct on at least this point: Natural language cannot bring into existence new representational powers that outstrip our nonlinguistic representational powers. I disagree, however, that our nonlinguistic thought exhausts all of the kinds of thoughts we are capable of thinking. I suggest a slight revision on the Fodorian dictum: Natural language cannot bring into existence new representational powers if the function of language is merely intrapersonal.

How might words increase the representational power of human cognition?

Wittgenstein once said: “Every sign by itself seems dead. What gives it life” (Wittgenstein 1953, ¶ 432). The question I am trying to push is, it seems to me, a related one: How might a mere word increase the representational power of human cognition? It seems to me that there are two different ways of attempting to give an answer here: Either words could capture details that cannot be captured by nonlinguistic representation, or words could make possible more abstract representations. Most of the attempts to provide a role for language in cognition that I have examined fall into the latter category. Thus, both Clark and Dennett see words as manipulable objects which stand for more complex (perhaps connectionist style) representations which are not themselves easily manipulable. The key here, however, as I have noted earlier, is that it does not seem that a word adds or takes away any information from the extant representation. Words can perhaps function as useful tools that stand for certain associations and eases the load on
working memory (as in Clark’s “scaffolded chimps”). But to admit this is not to admit that words in any way fundamentally alter our representational capacities. Neither Clark’s nor Dennett’s proposal gives us any clue to answering how words could increase the representational powers of human cognition. On my view this is not surprising since, having no special representational power of its own, the intrapersonal use of a word could not possibly increase the representational powers of nonlinguistic representation. This is not to deny, of course, that words may indeed act as either a bootstrap of scaffolding for memory-limited creatures such as ourselves. But in these cases the words are not actually increasing the representational powers of our cognitive systems.

Dennett and Clark view the power and “magic” of words in terms of a purely intrapersonal function: The function of words is purely internal to the individual’s cognitive processes. All of the other theorists I have engaged likewise confine their theories to the intrapersonal function of words (as I will show below). For the moment, however, I would like to try to pump the intuition that if the function of language is intrapersonal—if its role is, so to speak, only within the economy of the individual cognizer—then there is no hope of explaining how words increase the representational powers of human thought. Suppose that humans come with a rich store of prelinguistic conceptual representations. So, for example, a prelinguistic human can represent (as such) fruits and vegetables, bears and coyotes, etc. (We could suppose, further, that the human is an adult that as a child was raised by wolves and, as a result, has not acquired a natural language.) Suppose that we teach this “feral adult” words such as “fruit” and “vegetable” and “wolf” and “bear.” What new representational ability has this person now acquired as the result of learning to apply these words to the concepts he already
has? It is difficult to see how these words could somehow change his indigenous representational powers. If we furthermore suppose that his indigenous concepts are analogous to words in a language, then it becomes clear that it is no more comprehensible how learning the word “vegetable” could increase his representational powers than that an English speaker could change or increase her representational capacities by learning the German word “Gemüse.”

If words themselves are dead and have no representational powers of their own (which, as mere patterns of sound waves, they surely are), how could they possibly increase the representational powers of human thinking? I see no hope of providing any answer so long as one limits oneself to thinking about the intrapersonal function of words—that is, to the possible functions that words could play within the individual’s own cognitive system. If words are to increase representational powers, if they are to be given life, then I suggest we follow something like Wittgenstein’s own answer to his question about what gives signs life: Their life must come from their use. But to take this suggestion seriously requires that one look to the interpersonal function of words—that is, to the function that words have between individuals rather than within an individual.

The basic form of the argument I am making here is this: If language is to make a difference in the representational powers of thought, then it must be due to either the intrapersonal or the interpersonal function of language. I have suggested that it cannot be due to the intrapersonal function of language since we have no idea how a mere pattern of sounds (or some such pattern stored in memory) could increase the representational powers of human thought. Thus if language is to make a difference to the
representational powers of thought, it must be due to the interpersonal functions of language.

It remains a question whether language really does increase the representational power of human thought. Does language actually create new kinds of thinking which are in principle unavailable to nonlinguistic creatures? In particular, does it enable us to have thoughts whose contents are unavailable to nonlinguistic creatures? Many of the theorists I have engaged here seem to think that it does. I agree: language must be an important (if not the important) difference between the cognitive capacities of linguistic and nonlinguistic creatures. But this not something I am prepared or able to defend here.

It is interesting that none of the accounts I have looked at comes even close to countenancing the idea that the importance of language for cognition might come from its interpersonal functions rather than from its intrapersonal functions. As I have already noted, both Clark and Dennett seem to reduce the function of words to a purely intrapersonal one. Bermúdez likewise countenances only an intrapersonal function for language: making personal-level inferences. Carruthers follows suit, claiming that the function of language is to make certain kinds of thoughts accessible to conscious awareness. Carruthers’s claim that language makes possible cross modal thoughts is also an intrapersonal function of language since its role is wholly within the cognitive economy of the individual. Likewise, the psychologists Fei Xu and Jill de Villiers see language in terms of a purely intrapersonal function. Fei Xu thinks that words function as “essence placeholders,” leading the individual to construct a concept (the “essence”) by abstracting the properties that the words have in common. Words, for Xu, function as a certain kind of attention-grabber. Finally, Jill de Villiers thinks that the particular
syntactic structures of natural language make it possible to represent the truth of someone’s falsely believing something. De Villiers is a bit harder to pin down here since she wants to tell a story about how the use of language becomes internalized. For example, initially children recognize that sometimes people say things that aren’t true and then eventually are able to bootstrap from this to being able to attribute false beliefs to people. Nevertheless, if acquiring a syntactic structure (which then enables a new kind of thought) is supposed to be an interpersonal matter, that is something she doesn’t say.

I cannot show that the reason all these accounts fail is that they have not taken into account the interpersonal function of language, but I think the fact that all of the accounts limit themselves to the intrapersonal function of language, and the fact that all of the accounts fail to show that language really does play a role in cognition at least suggests that we might do better to look to an interpersonal account of the role of language in cognition. In what follows I will offer one way in which the interpersonal use of language might bring about new representational powers—and thus how the discontinuity thesis could be true.

**Discontinuity and the interpersonal function of language**

In the second chapter I introduced the idea of a similarity space: an n-dimensional space which enables a cognizer to see object A as more similar in some respect (according to various dimensions such as color, shape, texture, etc.) to object B than to object C. In chapter two I argued that these “similarity judgments” could account for the behaviors of Mandler’s infants. The similarity space does not represent kinds, since there are no boundaries which would delineate objects in different places in the similarity space as
different in kind. Objects can be more a less similar to one another in certain respects, but these similarities and differences are not sufficient to carve out kinds such as DOG or CHAIR. I will suppose that humans and many other animals have what I will call a *perceptual similarity space*: an n-dimensional space in which the dimensions are those to which the animal has some kind of perceptual sensitivity. The similarity spaces among members of the same species may be largely the same, but members of differing species may have radically different similarity spaces based on the kinds of phenomena that their perceptual hardware is tuned to. So, for example, a bat might have as one of the dimensions in its similarity space sonar profiles which represents the relative distances of different objects.

We can suppose that many nonlinguistic creatures have the ability to keep track of, say, different kinds of food by using their perceptual similarity space. For example, a creature that had eaten a mushroom in the past and had gotten sick might store that perceptual profile in its long term memory and avoid eating things in the future that were similar (that is, close to the stored exemplar in n-dimensional space) to it. Conversely, a creature that ate a mushroom that tasted good (and did not have any adverse side-effects) might attempt to eat mushrooms in the future that were similar to that good-tasting mushroom. Thus, by means of its perceptual similarity space, a creature could either choose or avoid certain kinds of mushrooms based on their perceptual appearance. So, for example, a creature that had eaten a mushroom that was white with a convex cap.

---

74 There is good neurophysiological reason to believe that rats can recognize situations (similar to those) in which they have been hurt in the past. Alcino Silva and others have shown that traumatic events such as a painful shock causes the release of CREB—a key chemical that accounts for the stability of synaptic structure and hence long term memory. Moreover, increasing CREB in transgenic mice had the effect that they had better memory of past fearful events (Kida et al., 2002).
(mushroom type 1) and had gotten sick as a result would recognize and avoid mushrooms with similar perceptual profiles in the future. On the other hand, a creature that had eaten a mushroom that was orange colored with an upside down cone-shaped cap (mushroom type 2) which had tasted good and didn’t have any adverse side-effects would choose to eat mushrooms that had a similar perceptual profile in the future.

In addition to a perceptual similarity space I suggest that linguistic creatures also have an *abstract similarity space*. Whereas the perceptual similarity space is determined by the values on the various dimensions to which the creature is perceptually sensitive, the abstract similarity space can represent more abstract properties such as “good” or “bad” (which are not directly perceptual properties). So, for example, a creature with an abstract similarity space might be able to group together objects in terms of their abstract properties such as “good” or “bad” despite the fact that objects within either category might differ wildly from each other. In addition, whereas the actions of creatures which have only a perceptual similarity space are driven by the perceptual similarity space, the actions of creatures that have, in addition, an abstract similarity space can also be driven by the abstract similarity space. So, for example, a creature with an abstract similarity space can choose to eat foods because they are “good” rather than because they are perceptually similar to foods that it has liked in the past. I suggest that such an abstract similarity space is reserved for linguistic creatures. I will try to show that language is needed for the existence of the abstract similarity space by considering how it might be formed.

How might we account for the genesis of an abstract similarity space? To return to our mushroom example, suppose that a new mushroom is found that has some of the
qualities of the poisonous mushroom and some of the qualities of the good mushrooms. For example, suppose that there is a mushroom that is white and has an upside down cone shaped cap (mushroom type 3). Since it shares some of the properties of the poisonous mushroom (white color) and some of the properties of the good mushroom (cone shaped cap) its perceptual profile will not enable one to make predictions regarding whether it is good or poisonous. Suppose, however, someone labels this new kind of mushroom with a word like “good.” And suppose that the orange mushroom with the cone shaped cap is also labeled with the same word “good.” The effect of labeling mushroom types two and three with the same word “good” could have the effect of creating a new abstract similarity space. Essentially the label would now make mushroom types two and three cluster closer together in (abstract) similarity because they would now share the property of being labeled with the word “good.”

This restructured similarity space is no longer a perceptual similarity space because its dimensions are no longer merely properties that objects have intrinsically; it is now sensitive to properties that humans create based on their various interests. Human interests are at play (e.g. avoiding poisonous mushrooms) and it is these, through the use of the word “good,” that have created a similarity space with a new dimension. This is the essence of what I am calling an “abstract similarity space.” The creature who has an abstract similarity space can now choose food not only based on perceptual similarity to food it has liked in the past, but can also choose food based on whether or not it is good.

In my imaginary scenario, the existence of the abstract similarity space makes it possible to choose the nonpoisonous mushrooms (even without having had to try them oneself). If we suppose that at least one person tries the new mushroom and doesn’t get sick, then the
word will spread that the mushroom is “good” and as a result people would see mushroom type 3 as more similar to mushroom type 2 than to mushroom type 1 since types 2 and 3 cluster together in (abstract) similarity space as a result of being associated with the word “good.”

Of course, there must be a story about how the word “good” comes to mean GOOD, and this is not a story I will not attempt to tell here. My point is just to suggest that words could play the fundamental role in creating an abstract similarity space that is structured by more than merely the perceptual properties of objects.

Is there any other way that an abstract similarity space could be created? Could perceptual similarity space restructure itself, for example? It is difficult to see how a perceptual similarity space could restructure itself since it can only capture the profiles of the perceptual properties that objects have. Smith and Heise 1992, who also invoke the idea of a similarity space which has some affinities to the one I am suggesting (i.e. an n-dimensional space where different objects have locations in that space), think that the similarity space can restructure itself. They propose that features in the environment that are correlated will be paid more attention, and that the result of paying more attention to these features’ weights these features more heavily in the similarity space. The result is supposed to be that those features which are weighted more heavily will cluster together in similarity space. So, for example, if beaks and wings are correlated, then these features will be weighted more heavily in the similarity space and as a result, the similarities and differences on these dimensions will have more of an effect on how objects with these properties will cluster. If so, then even though bats may be more similar in overall shape and color to crow than to flamingos, since beak and wing features
are weighted more heavily, the crow will actually end up clustering closer to the flamingo than the bat.

However, there are several problems with Smith and Heise’s idea (even aside from it being questionable whether we really do view crows as more similar to flamingos than bats). The main problem is that they seem to think that the dimensions that characterize a similarity space are things like “beak features” and “wing features.” But is this plausible? This seems to already assume the presence of concepts, which is what the similarity space (on their account) is suppose to explain. On my account, a perceptual similarity space is characterized by dimensions which all objects share and with which evolution could have plausibly endowed us. As I argued in chapter 2, the more specific we make our dimensions, the more likely it is that we are already presupposing concepts at least as complex as the ones we are trying to explain. This not to deny that we have the ability to recognize different nuances in beaks and wings, but only to say that there are not separate dimensions (wings and beaks) which characterize our similarity space. But if there is no separate dimension for wings and beaks, then our similarity space could not be restructured based on the correlations between those features since our similarity space would not be characterized by separate beak-dimensions and wing-dimensions. On the other hand, if the dimensions which characterize the similarity space are much more general dimensions such as color, shape and texture (as I have claimed), then it doesn’t really make sense to say that there are correlations. Consider: what sense does it make to say that there is a correlation between shaped-things and textured-things, or between colored-things and shaped-things? Thus, if the similarity space is characterized by the
kinds of dimensions I have suggested, then Smith and Heise’s proposal doesn’t even make sense, let alone work.

Even though Smith and Heise’s particular example doesn’t work, there could still be a sense in which perceptual similarity space could be restructured. It is possible that two segments of dimensions could be correlated in the sense that the number of points that fall within a certain segment of similarity space V on dimension X that also fall within segment S on dimension Y is a large proportion of the points that fall in either segment V or S. So, for example, suppose that most of the spherical shaped objects one encountered were also red. In that situation it would be the case that the “red” segment of the color dimension and the “spherical” segment of the shape dimension would be correlated in the sense that most of the points falling within either the red or purple segments of the similarity space also fall within the segment of similarity space which is the union of red segment and the spherical segment. If this were the case, then perhaps something like Smith and Heise’s idea could be made to work. That is, perhaps the creature would pay more attention to red spherical things than to, say, red cubed things. And perhaps the segments of perceptual similarity space would be weighted more heavily. However, it is difficult to see what good weighting a segment of similarity space would do (since, by hypothesis, most of the points are already clustered together there). Nor is it clear what it would mean to weight a segment of similarity space (as opposed to a dimension). But even if we could make good on a proposal of this kind, this kind of restructuring (where we do not already help ourselves to abstract concepts) doesn’t hold any promise for creating the kinds of abstract concepts we are interested in—concepts which are not based merely on perceptual similarity. So even if perceptual similarity
space could be restructured in certain ways, these ways will not get us to an abstract similarity space. Rather, it seems that the only way to create an abstract similarity space—one not driven merely by perceptual properties—is through the use of language.

It seems, then, that a natural language is required for the existence of an abstract similarity space, and that there is no clear way that it could be created by anything other than language. In particular, it couldn’t be created by perceptual similarity space restructuring itself. If this is correct, then this (admittedly underdeveloped) account of how an abstract similarity space might be created is an avenue to pursue in attempting to vindicate the discontinuity\textsubscript{L} thesis. The abstract similarity space increases the representational power over the merely perceptual similarity space since the way things are represented within the abstract similarity space differs from the way the perceptual similarity space would represent them. It is an increase since creatures that have an abstract similarity space retain their perceptual similarity space and so the mushrooms can still be represented as more or less similar to each other with respect to their perceptual properties. So for example, although two different kinds of mushrooms may differ significantly in their “perceptual profiles” (i.e. types of mushrooms whose placement in perceptual similarity space differs in that the different types of mushrooms cluster in different areas of the perceptual similarity space) they may nevertheless cluster together in the abstract similarity space in virtue of both being instances of “good” mushrooms (i.e. mushrooms which are associated with the word “good”). To say that the abstract similarity space increases the representational power of cognition and that this cannot be done by the perceptual similarity space restructuring itself is to make good on the discontinuity\textsubscript{L} thesis since the key idea of the discontinuity\textsubscript{L} thesis is that there is a
kind of cognition which language makes possible and which differs in the kinds of representational contents that are possible (i.e. the abstract similarity space).

Importantly, the account that I’ve offered here is one according to which the function of language in creating the abstract similarity space is interpersonal rather than intrapersonal. It is because of the way that language is used among the speakers of the community that the abstract similarity space is created and maintained. Suppose that I am a neophyte who has not yet had an abstract similarity space created with respect to mushrooms. I cannot yet tell a good mushroom from a poisonous mushroom although I can see certain mushrooms as more or less similar based on their position within my perceptual similarity space. Suppose now someone calls certain kinds of mushrooms “good.” These kinds of mushrooms that I see people call good now have a new property that makes them distinct from other mushrooms that do not have the word “good” attached to them (or, alternately, have a different word like “poisonous” attached to them). I am not here supposing that I have kind concepts for the different kind of mushrooms. Rather, they cluster together in virtue of their similar perceptual properties. When someone applies the word “good” to mushrooms of a certain type, I will have to learn to apply that word to mushrooms of the same type, but I do it by seeing a mushroom as similar to the ones that I have seen “good” applied to in the past. I have suggested that the use of the word has the effect of making different kinds of mushrooms (which are associated with the same word) similar to each other (within the abstract similarity space) despite their other perceptual differences. But importantly, the abstract similarity space depends on the interpersonal use within a language community. The importance of language here is how it functions among speakers; I could not form an
abstract similarity space in this case if it were not for other speakers. As a neophyte I depend on other speakers to shape my abstract similarity space and bring it into conformity with the abstract similarity space of other speakers in my community. So before my abstract similarity space has been formed I could sort mushrooms in various ways based on the perceptual properties they shared, whereas after my abstract similarity space has been formed I can sort them into good and poisonous. The process of forming an abstract similarity space will require correction from other members of my community. If we are out gathering mushrooms and I (a neophyte) attempt to put a type I mushroom into the basket, I will be corrected by another member of the group.

There is another important sense in which the interpersonal function of language is key for the existence of the abstract similarity space. Whereas the perceptual similarity space is something that individual cognizers have, the abstract similarity space is not the unique possession of an individual. One could not have an abstract similarity space if one didn’t first have a perceptual similarity space since we use our perceptual similarity space in acquiring an abstract similarity space. For example, if I could not distinguish certain types of mushrooms by their various properties, then I would not be able to progress on to having an abstract similarity space. However, there is an important sense in which the abstract similarity space is not merely something that I have (as the perceptual similarity space surely is). Rather, its existence is ultimately upheld by the linguistic norms of the community. Thus there is an important sense in which the abstract similarity space is not something the individual has all to herself. Of course, it is individuals who benefit from and make use of the abstract similarity space, and it is
individuals who uphold the linguistic norms. But the abstract similarity space is something that no one individual can create and maintain.

Although I will not develop the idea in this dissertation, with the advent of the abstract similarity space (which itself is created and upheld through linguistic norms—such as when it is appropriate to apply “good” to a mushroom) comes a new kind of representation: conceptual representation. Kind concepts are not to be identified with any cluster in similarity space. Rather, inasmuch as cognition involves the classification of things into kinds, that job could be performed by linguistic labeling. That is, the discontinuity between conceptual and nonconceptual thought is the same discontinuity as that between linguistic and prelinguistic thought: $\text{Discontinuity}_L$ and $\text{Discontinuity}_C$ map onto each other. Another way to put the point is: the semantics of prelinguistic thought (perceptual similarity space) differs radically from the semantics of linguistic thought (linguistic labeling, linguistic norms). I have not defended anything like this in this dissertation, but it could plausibly be seen as a further development of the alternative way of defending the discontinuity $L$ I have suggested. If, as I have suggested, linguistic thought is discontinuous from prelinguistic thought, then one would like to know about the semantics of these different types of thought. This remains an important and highly contentious project within the interstices of philosophy of mind and language.

What I do take myself to have done in this chapter is to motivate an alternative conception of the role of language in cognitive development. I have argued that most of the theorists I have engaged in the course of this dissertation all end up embracing some form of the continuity $L$ thesis—hence denying that language makes a difference to the representational power of human thought. Furthermore, I have noted that all the theorists
have limited themselves to viewing the function of language in cognition as merely intrapersonal and that this limitation may be a clue to why their accounts fail to account for a difference in representational power between linguistic and nonlinguistic thought—or at least a motivation for looking to a different kind of account. I also gave a reason to think that language just couldn’t make a difference to our representational capacities if the function of language is viewed as intrapersonal and subjective. Finally, I have suggested a way to vindicate the discontinuity hypothesis by presenting a way in which the function of language might be interpersonal, external to the individual and irreducibly social. The idea of an abstract similarity space leaves many questions, such as the semantics of linguistic thought, unresolved. Nevertheless, it helps make the point that there are some kinds of thoughts that may be unique to linguistic creatures inasmuch as there are some properties that cannot be represented using the creature’s natural endowments (such as a perceptual similarity space).
REFERENCES


De Villiers, Jill. 2005. Can Language Acquisition give Children a Point of View? In Astington, J. and Baird J.
Dewar, K. and Fei Xu. Unpublished manuscript. Do 9-month-old infants expect distinct words to refer to kinds? (http://www.psych.ubc.ca/~fei/)
http://ruccs.rutgers.edu/pub/papers/spelkems.pdf


Montgomery, Derek E. 2005. The developmental origins of meaning for mental terms. In Astington, J. and Baird, J.


Perner, Josef, P. Zauner, and M. Sprung. 2005. What does “that” have to do with point of View? Conflicting desires and “want” in German. In Astington, J. and Baird J.

Perner, Josef, M. Sprung, P. Zauner, and H. Haider. 2003. Want That is Understood Well before Say That, Think That, and False Belief: A Test of de Villiers’s
Linguistic Determinism on German-Speaking Children. *Child Development*, 74, 179-188.


*Cognitive Psychology*, 30, 111-153.

Xu, Fei.  1997.  From Lot’s wife to a pillar of salt: evidence that physical object is a 

Xu, Fei and Susan Carey.  2000.  The emergence of kind concepts: a rejoinder to 

Xu, Fei.  2002.  The role of language in acquiring object kind concepts in infancy. 
*Cognition*, 85, 223-250.

Xu, Fei.  2003.  Numerosity discrimination in infants: Evidence for two systems of 

Xu, Fei, Susan Carey, and N. Quint.  2004.  The emergence of kind-based object 

Gershkoff-Stowe and D. Rakison (eds.), *Building object categories in 
developmental time*.  Papers from the 32nd Carnegie Symposium on Cognition. 
New Jersey: Lawrence Erlbaum.

Xu, Fei.  In press.  Concept formation and language development: count nouns and object 
University Press.  (http://www.psych.ubc.ca/~fei/)

Xu, Fei, Cote, M. & Baker, A.  2005.  Labeling guides object individuation in 12-month-
old infants.  *Psychological Science.*