Core Cognition and its Aftermath

Lance J. Rips

Northwestern University

Thanks to Erin Anderson, Jacob Dink, Sandy LaTourrette, Nick Leonard, Alexis Wellwood, and especially Susan Hespos for coaching me on these topics. Thanks also to audiences at Northwestern University, the International Conference on Thinking, and the Conference on Epistemology and Cognition (particularly to commentators Anna-Sara Malmgren and Ram Neta) for very helpful comments.
Here’s a safe assumption about cognition (if anything is): Infants innately possess perceptual skills that they can later exploit in learning. But much of what these infants have to learn seems to be non-perceptual. For example, children eventually have to master the ideas of numbers and of causal relations, both of which seem abstract and impossible to perceive. Although we can see numerals like “6” and “VI,” we think of both these numerals as representing something more abstract—the number six. And although we can see one physical object making contact with another, we may not be able to see the causal connection between them. We can also think about causes and effects that take place among things that are too small or too remote in time or that are simply invisible—interactions among subatomic particles, prehistoric glacier movements, or orbits between pairs of black holes.

If innate perceptual abilities deliver information that’s too low-level, they can’t by themselves explain how children gain abstract concepts. If perception only provides information about the bare location of particular colors in the visual field, for example, then there’s little hope that children could use such facts to learn their way to ideas about number and cause. It’s even hard to see how they could gain the idea of an ordinary physical object, like a glass or a tree (as distinct from glass or wood), from such spare beginnings. Of course, parents and teachers provide them with further information about abstract and concrete individuated entities, but in order for the children to profit from these lessons, they need suitable background concepts. Otherwise, children would be in the position of an adult trying to learn a language from scratch by consulting a dictionary and a grammar written entirely in that language. Learning would be impossible because they couldn’t get a toehold.

To take up the slack between innate perceptual abilities and abstract concepts, developmental psychologists have at least two choices. They can assume that, in addition to innate skills tied to perception, people also possess innate non-perceptual concepts that they can combine with the perceptually bound ones. Perhaps people have innate non-perceptual concepts of physical object, causality, and number that they can harness to understand these domains in the way adults do. The alternative possibility is to sophisticate the perceptual abilities, to think of them as complex enough to serve as building blocks for later adult ideas. Perhaps children can grasp objects, causality, and numbers by constructing these concepts from the sophisticated initial ones.
This article looks at one version of the second strategy that goes by the name “core knowledge” or “core cognition.” (“Core knowledge” is the usual term, but I’ll stick to “core cognition” here to avoid confusion with epistemologists’ use of “knowledge.”) The core-cognition program assumes that infants have innate perception-initiated abilities that provide them with concepts such as individuated physical object, number, and cause. Although these concepts fall short of those of adults, nevertheless children can combine the core concepts in a way that eventually yields the adult ideas. Later sections of this article describe some of the evidence that supports core cognition. I’ll be assuming that core abilities do exist in each of the domains I’ll examine. The focus will instead be on the implications of core cognition for later learning.

I’ll be arguing that children can’t arrive at an adult understanding of these domains just by combining core information. In each case, something is missing from the core—probably innate nonperceptual content. In a little more detail: According to the core-cognition theory, a core system receives perceptual input but is sealed off from information coming from other core systems and from central cognition. It’s encapsulated, in a sense I’ll review in Section 1. Second, although a core system might be capable of some forms of learning, it doesn’t transform itself internally in a way that produces the full set of adult abilities in the domains of interest here. This is an empirical claim, based on current findings, discussed in Sections 2-4. For example, none of the proposed core systems for number can develop internally into a system that’s capable of adult arithmetic. Together, encapsulation and limited learning put a cap on how sophisticated a single core system can be. So far, these points are common ground with most proponents of core cognition. But I’ll also be claiming that children can’t derive the key adult concepts by combining different core systems. So the building-block conception of core cognition can’t be correct. Core systems don’t provide mental components that are sufficient to explain adult concepts.

That’s the story line, but first a bit more about core cognition itself.

---

2 The distinction at stake here is not between perceptual and conceptual information, but between systems whose input is perceptual (and which are cutoff from central cognition) and the central systems themselves. Both core and central systems include concepts—mental representations that have referents and that figure in inferences. Sections 1 and 5 of this article attempt to clarify the relevant difference.
1. Core Cognition

The idea of core cognition comes most directly from Elizabeth Spelke (e.g., Spelke, 2000, 2003; Spelke & Kinzler, 2007), although many other developmental psychologists have influenced and been influenced by it (see, e.g., Carey, 2009, and the recent collection by Barner & Baron, 2016). However, the origins of core cognition trace back to Fodor’s (1983) well-known theory of perceptual modules.

Perceptual modules are devices that are specialized for dealing with particular kinds of input. As soon as a module picks up the information for which it’s specialized, it automatically processes that information. The module typically includes processing steps that transform the input from the original sensory data to more complex mental representations. Then, at the end of this series of transformations, the module outputs the representation in a form that central cognition can use. One example of a module is the perceptual system that figures out the 3D shape of an object from its 2D image. A three-dimensional object like a cup projects a two-dimensional image on your retina. The perceptual module for 3D shape takes that retinal information, passes it through a series of computations, and reconstructs a mental representation of the 3D cup that can then be used by other cognitive processes.

Modules run independently of higher-level cognition. Modules are “informationally encapsulated,” to use the term of art. That means that your expectations about what you will see and your memories of what you have seen don’t affect the modules’ operations. It’s only at the very last stage, after the module’s representation has been kicked upstairs to central cognition, that your beliefs have access to it. In addition, perceptual modules are dedicated to just one kind of task, such as dealing with 3D shape. In other words, they’re task-specific. Perceptual modules also operate in a mandatory way. For example, you can’t keep yourself from seeing a 3D object as 3D. Furthermore, they finish their task rapidly, so rapidly that you’re not aware they’re at work. And they’re innate; they’re hardwired.

Most of these properties of perceptual modules find their counterparts in core cognition. Spelke (2000) explicitly mentions encapsulation and task specificity as among its properties. They are also innate (Spelke, 1994); the elements of core cognition have “a long ontogenetic and phylogenetic
history” and “are very similar to those of many nonhuman animals” (Spelke, 2000, p. 1233). For that reason, core cognition is universal among humans (Hespos, 2016; Spelke, 1994): “They are a body of knowledge that all humans share, whatever the diversity of our elaborated belief systems” (Spelke, 1994, p. 441). By way of a direct comparison to Fodor, Spelke (2003, p. 31) remarks that for the core cognition systems of objects, number, agency, and places:

Studies of non-human animals and human children suggest that each of these systems is a cognitive module in Fodor’s sense (Fodor 1983), that each is largely constant over human phylogeny and ontogeny, and that together these systems provide the building blocks of mature human intelligence.

The main difference between core cognition and perceptual modules is one of degree of integration. Spelke’s comments on possible differences between these theories are brief (see Spelke, Breinlinger, Macomber, & Jacobson, 1992, p. 605), but core cognition seems to pool processes that are likely to be independent modules according to Fodor’s account. For example, a mechanism for computing the 3D shape of an object and one for computing its color are probably separate Fodorian modules, but parts of a single core component for the analysis of physical objects (see Xu, 2016, for a similar point). One advantage of integration is that it allows a core system to combine information from different sensory modalities. For example, a single core system for number can keep track of both the number of visually perceived objects and the number of auditorily presented tones.

Two other properties of core systems seem to follow from the idea that these systems are “building blocks of mature human intelligence.” First, core cognition continues to exist in adults. These systems are permanent, never overwritten by later processes. They appear overtly in adults’ performance, sometimes when other processes are unavailable due to time limits or other restrictions. The building-blocks approach requires that core representations continue to exist since they function as components of more complicated cognitive operations. (“Permanent,” in this sense, doesn’t imply unchangeable. Refinements in a core system can take place through maturation; see Spelke, 1998.) Second, and most important for the present discussion, the output of one core ability can combine
with others to explain complex, adult-type thinking. Although a single core ability can’t directly talk to another (because they’re encapsulated), they can combine in central cognition by means of natural language or associative learning to produce complex ideas.

Most proponents of core cognition ascribe to it the properties just mentioned. However, one salient difference among proponents is whether they believe these abilities are sufficient to explain adult concepts. Carey (2009), for example, believes that concepts of the integers require radical conceptual change: These concepts can’t be spelled out in the vocabulary of the core systems for number.³ By contrast, Spelke tends to downplay the existence of such radical changes in development in general (e.g., Spelke et al., 1992) and in number learning in particular (Spelke, 2011a). This article focuses on Spelke’s building-blocks approach in order to examine, in its pure form, the idea that core cognition is the foundation for adult thinking. I’ll be arguing that Carey is right that core cognition is not always sufficient to explain key aspects of adult concepts; for example, it’s not sufficient for the concepts we’re about to look at.⁴

In the next three sections of this article, I’d like to proceed inductively by examining the role core cognition is supposed to play in people’s thinking about mathematics, causality, and referents of count and mass nouns. These examples should provide us with a more concrete picture of how core cognition works. Within each domain, I’ll try to describe the initial core abilities and the way the theory uses them to explain the development of more advanced thinking. The goal will be to identify the difficulties that core cognition encounters in explaining how adult concepts emerge from their core basis. Of course, I won’t be able to prove that core cognition is unable to account for adult

---

³ A theory like Carey’s (2009) could consistently maintain that new adult representations replace the infants’ core representations. Conceptual change could produce concepts that make core cognition unnecessary. For example, the adult concept of the integers could replace its core predecessors. This option presumably isn’t possible for Spelke, since she needs core representations as foundations for adult thinking. However, Carey (2009, pp. 67-68) believes that core cognition does continue to exist in adulthood, on the basis of empirical evidence that suggests its presence (e.g., Kahneman, Treisman, & Gibbs, 1992, in the case of the object system; Barth, Kanwisher, & Spelke, 2003, in the case of the number system; see Sections 2 and 3 for discussions of these systems). As hinted a moment ago, core systems may continue to play useful roles in adult thinking, sometimes serving as fallback procedures when more sophisticated ones are temporarily unavailable.

⁴ I agree, however, with Spelke’s doubts about Carey’s theory of radical conceptual change. But I’ve discussed this issue elsewhere—see Rips, Asmuth, and Bloomfield (2006, 2008); Rips, Bloomfield, and Asmuth (2008); Rips and Hespos (2011)—as have others (e.g., Fodor, 2010; Rey, 2014), and I won’t rehearse these worries here.
thinking. Proponents can always complicate existing core systems and can posit new ones. But I’ll try to argue that inherent features of core cognition stand in the way of an adequate theory.

2. Core Cognition and the Count/Mass Distinction

As a first example, let’s consider a recent paper by Brent Strickland, called “Language Reflects Core Cognition” (Strickland, 2016). This paper tries to make the case that core cognition can explain some common linguistic features, such as the count/mass distinction, causative constructions, gender marking, and thematic roles. I’ll focus on the count/mass distinction here, since it’s Strickland’s best-worked-out example.

In languages such as English, German, Italian, and Russian—number-marking languages—the mass/count grammatical distinction appears in the way nouns combine with plural markers, determiners, and quantifiers (e.g., Chierchia, 2010; Gillon, 2012; Pelletier & Schubert, 2003; Rothstein, 2010). Mass nouns, such as mud and gold, can’t take plural forms. You can’t say *muds or *golds without a change in meaning. Similarly, mass nouns can’t be used with numeric quantifiers—*five mud(s) or *ninety-nine gold(s) are ungrammatical. Count nouns, like robin and toaster, have the opposite characteristics: robins, five robins, toasters, and ninety-nine toasters are all perfectly fine. Not all languages distinguish count and mass nouns. Classifier languages, including East Asian languages, such as Chinese and Japanese, and South American languages, such as Arawak and Nahuatl, don’t have mass/count marking on nouns (see Aikhenvald, 2000, for a survey). However, the conceptual distinctions associated with mass/count may surface in these languages in other grammatical forms.

Now how is core cognition supposed to explain the mass/count distinction? As Strickland (2016) points out, research on infants suggests that even 5-month-olds differentiate solid objects like an ice cube from non-solid stuff like liquid water. Figure 1 illustrates the procedure from one such experiment (Hespos, Ferry, & Rips, 2009). In this study, 5-month-olds saw a clear cup with some blue stuff inside. During the first part of the procedure—the habituation sequence—the experimenter tipped the cup and rotated it on its bottom edge until the infants lost interest and looked away. For one group of infants, the blue stuff was liquid (upper left part of Figure 1), and the top surface of the
Core Cognition and its Aftermath / 8

liquid stayed parallel with the surface of the table as the cup rotated. For the other group of infants, the blue stuff was a solid plug of plastic that was created to look exactly like the liquid when it was upright and stationary. But when the cup rotated in this condition during habituation (upper right part of Figure 1), the top of the solid was always parallel to the bottom of the cup, rather than to the surface of the table. Finally, we tested all the infants by showing them two different scenes, one-at-a-time (bottom panels of Figure 1). In the solid test trials, the experimenter tipped the contents of the cup into another cup, and the infant saw the plug of plastic tumbling out. In the liquid test trials, the experimenter again tipped the contents, but the infant saw the blue liquid pouring out.

The prediction is that if the infants in the liquid condition had figured out during the habituation sequence that the contents of the cup was liquid, their attention should be aroused during the test trials if the contents tumbled out rather than poured out. Researchers standardly assume that greater attention (or surprise) motivates infants to look longer in these procedures. So infants in the liquid condition should look longer at the tumbling than at the pouring. And in the same way, if the infants in the solid condition had figured out that the contents of the cup was solid, they should look longer during the test trials if the contents poured out rather than tumbled out. That’s in fact what the results showed. Figure 2 plots the average amount of time that the infants looked at the display during the test trials, as a function of (a) whether the infants were habituated to the liquid or the solid, and (b) whether the test display was a pouring or a tumbling. The figure shows the predicted crossover: Infants looked longer at the unfamiliar type of item than at the type they had seen during the habituation sequence. Five-month-olds seem to recognize the difference between solids and liquids well enough to be able to predict what they will do under further transformations. Hespos et al. (2009) also showed that after habituation to a liquid, infants expect that a solid cylinder will be able to penetrate the surface of the liquid and will be surprised if it remains resting on the surface. But if they are habituated to a solid, they expect the cylinder to rest on top of the surface of the solid and will be surprised if it goes through (see, also, Hespos, Ferry, Anderson, Hollenbeck, & Rips, 2016, for similar results for other nonsolid substances).

As Strickland (2016) mentions, results like these suggest that infants’ core equipment can distinguish solid objects from nonsolid stuff. What’s important in the present context is whether this
core distinction could help explain the grammatical count/mass distinction. According to Strickland, core cognition of this sort makes the difference between solid objects and nonsolid stuff more likely to come up in conversation among adults, leading languages over time to grammaticalize the difference as count versus mass syntax. Children learning this language will, in turn, find the grammatical distinction easy to learn because it corresponds to core cognition, and ease of learning will help perpetuate the count/mass marking in the grammar.

A difficulty with this proposal, however, is that the distinction that infants are attending to in experiments like the one I just described is quite different from the count/mass distinction in English and other languages. The infants are presumably reacting to the solidity or nonsolidity of the contents of the cup in Hespos et al.’s (2009) study. The solidity of the blue plug is what keeps the contents stable as the cup rotates and makes the plug tumble out when it’s tipped. The nonsolidity of the blue liquid is what makes it move about as the cup rotates and pour out when the cup is tipped. But solidity and nonsolidity don’t line up especially well with the count/mass distinction. Table 1 maps these categories and illustrates some counterexamples as dashed lines that cross over from solids to mass nouns or from nonsolids to count nouns. One misalignment is due to the fact that solids can be named with either count nouns like toaster or with mass nouns like gold, wood, iron, brass, and many others. Similarly, nonsolids can be named with mass nouns like water, but also with count nouns like soufflés. Second, as Strickland mentions, some solid objects like cows or coins can be named either by count nouns, like the ones I just used, or by mass nouns, such as cattle or change. There are also many so-called “fake mass nouns” or “object mass nouns,” such as jewelry, silverware, furniture, clothing, and fruit, that are mass nouns but name solid things. Then, third, both count nouns (e.g., suggestions) and mass nouns (e.g., advice) can refer to abstract entities, which are neither solid nor nonsolid. So the mapping between the distinctions seems quite complex. These issues are all well known in linguistics and philosophy of language (see the sources mentioned in the second paragraph of this section).

However, we need to be careful in comparing Strickland’s (2016) theory to the evidence just presented. Strickland seems to interpret earlier results (e.g., Cheries, Mitroff, Wynn, & Scholl, 2008) to show that core cognition latches on to a particular type of object (sometimes called a “Spelke
Object (Spelke object) that consists of a spatially connected region that moves as a whole against a stationary visual background. Cats, toasters, and other middle-sized movables are Spelke objects in good standing. Other non-Spelke-object entities are treated as a residual class of unspecified stuff that, as Strickland (2016, p. 7) says, “often but not necessarily encompass substances.” The objects of core cognition (Spelke objects) again get mapped to count nouns. The remaining non-objects can be named in a more flexible way by either count or mass nouns. Hespos et al. (2009, 2016) argue, to the contrary, that a separate domain of core cognition is responsible for infants’ ideas about nonsolids, like water and sand, as in Table 1; infants don’t simply treat them as the absence of Spelke objects. But let’s consider the Spelke-object/nonobject split in order to get closer to the spirit of Strickland’s proposal. A better rendition of this theory might therefore be the set of contrasts in Table 2. This revision may allow us to get around some of the messiness of Table 1 if we assume that solid substances like gold are treated as nonobjects and that nonsolids like soufflés can be objects.

Of course, we’re still faced with clear-cut objects, like cattle/cows, that are lexicalized as either count or mass nouns. We could say, as Strickland hints, that mass nouns are nonselective and can name either objects (e.g., “cattle”) or nonobjects (e.g., “water”) (see Barner & Snedeker, 2005, and Srinivasan & Barner, 2016, for a theory of this sort). But in addition, many objects that are named by count nouns are not Spelke objects. These include, not just abstract entities like suggestions and ideas, but also ordinary concrete entities such as trees and shrubs, houses and other buildings, floors and ceiling, sidewalks and roads, lakes and rivers, noses and ears, handles and rims (of cups), and many other things that don’t move independently. So nonobjects can map to both count nouns (e.g.,

---

5 Yet another way of connecting core cognition with the mass/count distinction appears in Odic (2014). According to this view, count nouns recruit one core system—the approximate number system—for purposes of quantification (see Section 3 for more on this system). A question like Are more of the blobs yellow or blue? is then answered by comparing number of blobs. Mass nouns, however, recruit a different core system—the approximate area system—for quantification. A question like Is more of the blob yellow or blue? is answered by comparing areas. Odic reports differences in the accuracy of people’s answers to such questions (even when stimulus properties are controlled), suggesting the use of the two systems. All accounts of core cognition distinguish the object system from the approximate number system; so the distinction between approximate number and approximate area is quite different from either of the core distinctions that appear in Tables 1 and 2: It’s not equivalent either to the solid/nonsolid split or to the Spelke-object/nonobject split. However, Odic (2014) may not be intending a rival account to those of Tables 1 and 2. That is, he may be proposing these systems, not as the cognitive basis of the linguistic count/mass difference (in the way that Strickland does), but instead as a reflection of the structure that these noun phrases independently impose. For this reason, I won’t pursue the theory here.
“houses”) and mass nouns (e.g., “water”), and mass nouns can encode both objects (“furniture”) and nonobjects (“water”), as shown in Table 2. We’ve again lost the straight-through mapping of objects to count nouns and of nonobjects to mass nouns, as the dashed arrows in Table 2 show. Neither description of core cognition provides a necessary and sufficient basis for the count/mass distinction (see Rothstein, 2010, for a similar conclusion).

This raises the problem of how we get from the core cognition distinction—whether solid/nonsolid or Spelke object/nonobject—to the grammatical count/mass distinction as it appears in language. Core cognition was supposed to explain this grammatical difference, but so far we have no way around the many exceptions to its predictions. This problem wouldn’t be so bad if the exceptions to a straightforward mapping were idiosyncratic (i.e., language specific) lexical items. Perhaps cases like cows/cattle, coins/change, pots/pottery, and object mass nouns (e.g., jewelry, clothing, furniture) qualify as idiosyncratic (but see Grimm & Levin, 2012, and Wisniewski, Lamb, & Middleton, 2003). But other exceptions seem systematic. If we take core modules for solids versus non-solids as basic (as in Table 1), we have to deal with a seemingly open class of solid substances that get lexicalized as mass nouns (e.g., wood, iron, gold, ore, plastic, …). If we opt instead for a Spelke-object module (as in Table 2), we have to contend with a similarly open-ended set of items that don’t move on their own (so are non-Spelke-objects) that are lexicalized as count nouns (e.g., buildings, roads, mountains, ears, …). In other words, being a Spelke object seems to be neither necessary for count-noun status (since count nouns like trees name non-Spelke-objects) nor sufficient for count-noun status (since Spelke-objects like cows can be named with by mass nouns like cattle). At the very least, then, we have a gap in the explanation that is supposed to take us from Spelke objects to count nouns.

What I’d like to claim is that this problem of getting from core cognition to the relevant adult concepts is symptomatic. Even if we grant the usual interpretation of what kind of information core cognition is detecting, we don’t get a clear learning story about how people go from the innate systems to the adult constructions. Let’s consider a second example to get a different perspective on this problem.
3. Core Cognition and the Positive Integers

Proponents of core cognition have also used it to explain infants’ number-related abilities and the way these abilities support later learning of the positive integers. This topic has provoked much more debate than the core-cognition approach to the mass/count distinction, and several versions of this approach are on the table (e.g., Carey, 2009; Piantadosi, Tenenbaum, & Goodman, 2012; Spelke, 2000, 2011a, 2011b). For present purposes, I’ll focus on Spelke’s theory because it gives a more prominent role to core cognition than its rivals.6

Spelke’s theory begins with the idea that infants have two number-relevant systems of core cognition. One of these systems we’ve already met: It’s the Spelke-object system that we glimpsed in connection with count nouns in Table 2. For purposes of explaining number knowledge, what’s important about this system (as it exists in infants) is that it can keep track of up to three objects simultaneously as these objects move around the visual field. The system doesn’t directly represent the number of objects it is tracking. It just represents the objects (e.g., a pencil at a particular location). But the system is relevant to number in that if one of the objects it’s tracking goes missing, then it will notice this fact. And if a new object comes into view, the system will also notice the change, as long as the total number of objects in the scene is less than four.

The second relevant piece of core cognition, in Spelke’s account, is a system that does directly represent numerosity or cardinality. This system goes by several aliases, like number sense or mental number line or analog magnitude system or large number system, but most people these days call it the approximate number system or ANS; so I will too. The ANS is sensitive to the total number of items that a person is looking at, and it represents that total as a single continuous quantity. You can imagine that people have some reservoir of mental activation that increases continuously with the number of objects on display, so that the resulting amount of activation provides a rough index of the total number of objects. However, the record the ANS keeps of the number of objects is not a linear function of the actual number. Like most relations between physical quantities and their mental

---

6 For comments on Carey’s (2009) proposal, see Rey (2014); Rips, Asmuth, and Bloomfield (2006); and Rips, Bloomfield, and Asmuth (2008); for comments on Piantadosi et al. (2012), see Rips, Asmuth, and Bloomfield, (2013).
representations, the ANS is more sensitive to differences between small quantities than to (objectively equal) differences between large ones. For example, the ANS represents the difference between four and five objects as bigger than the difference between fourteen and fifteen objects. The exact format of the ANS’s representation is a matter of debate (see, e.g., Cantlon, Cordes, Libertus, & Brannon, 2009, vs. Dehaene, Izard, Spelke, & Pica, 2008). But the response of the ANS is some bending-over function of the actual number of objects, like a logarithmic function, and I’ll refer to it as “log-like” from here on.

Support for the existence of the ANS in infants comes from an experiment that habituated six-month olds either to displays containing 8 dots or to displays containing 16 (Xu & Spelke, 2000). Figure 3 shows the nature of these configurations. Infants in one group saw a sequence of 8-dot displays (left side of the figure). A second group of infants viewed a sequence of 16-dot displays (right side of the figure). These sequences varied the size of the dots to control for variables such as the total area taken up by the dots. After habituation, both groups were tested by showing them new displays of 8 and 16 dots (bottom of Figure 3). The prediction is that if infants are able to register the number of dots during the habituation sequence, then those habituated to 8 dots should be surprised to see 16 at test but not to see yet another batch of 8. Similarly, infants habituated to 16 dots should be surprised to see 8 at test but not 16. In other words, they should look longer at the new number of dots than at the old number. And, in fact, mean looking times were reliably longer for the new number. Thus, six-month olds seem able to discriminate cardinalities larger than what could be handled by their object-tracking system, which is limited to three items.

Now, back to the question of how people get from these core systems to knowledge of the positive integers. Most proponents of core cognition believe that neither the object system nor the approximate number system alone can produce this knowledge. You can’t do much with the integers if you only have access to representations of one, two, or three objects, which is the range of the object system. And in the same way, when you get to first grade, if you still think that the difference between four and five is larger than the difference between fourteen and fifteen, as the approximate number system dictates, then you are going to get a very low grade. Neither system has the right properties to capture the integers. So how do we learn the correct integer properties?
According to Spelke’s (2000, 2011a, 2011b) proposal, grasp of the correct integer properties comes about as two-to-four-year old children learn the meaning of the number words by combining the properties of the object system with those of the approximate number system. Once children understand that the words “one,” “two,” and “three” map to both systems, they can figure out that “one” refers to a cardinality of approximately one and also that the same word refers to a set containing a single object. Similarly, the word “two” refers to a cardinality of approximately two and to a set containing two objects. The same goes for “three.” Given this information, they can then work out the fact that the addition of one individual to a set is coordinated with an increase in approximate cardinality. This is supposed to get around the limits of both systems and to produce a correct representation of the integers. As Spelke (2000, p. 1238) puts it:

The language of number words and the counting routine allow young children to combine their representations of objects as enduring individuals with their representations of numerosities to construct a new system of knowledge of number, in which each distinct number picks out a set of individuals with a distinct cardinal value…More specifically, the object system is the source of the child’s understanding that number applies to discrete individuals and that numbers can be changed by adding one, and the approximate numerosity system is the source of the child’s understanding that number applies to sets and that sets can be compared according to cardinal value.

You might be able to put yourself in the shoes of children trying to learn the integers in this way by thinking about your perceptual impressions of large discrete groups, like the number of jellybeans in a jar or the number of people at a concert. You know you are dealing with a group of individuals, and you also have a feel for the approximate number of items (though not their exact number)—maybe a few hundred. Children learning the integers in Spelke’s way have the same two sources of information associated with the number words “one,” “two,” and “three,” when applied to a group of objects.
But how do the children get from these two core-cognition properties to knowledge of the correct properties of integers? Spelke (2011a, p. 149) says that “Without regard to the order of these [number] words, children could discover how these…representations relate to one another: progressively larger cardinal values result from progressive addition of one.” In interpreting Spelke’s comment (and the one quoted earlier), we have to be careful not to read too much into “addition of one,” since children of three or four who are just learning the first few integers don’t understand addition of one in the mathematical sense of +1. So “addition of one” must mean something simpler, like placing a single object together with a group of other objects. The theory would seem to be that when kids group objects in this one-by-one way, they get an impression of larger cardinal values, as signaled by their approximate number system. They can then induce the rule that for each addition of one there is a further increase in cardinal value.  

Remember, though, that this impression of larger cardinal values is the continuous log-like impression of the ANS. Adding one to a group of none produces a much larger impression of the difference than adding one to a group of one, and that impression is greater than the difference produced by adding one to a group of two. So it is unclear how combining core representations in this way is able to produce a correct understanding of the integers. What children eventually need to know in order to understand the positive integers is that there is a first number (1 for the positive integers), that every positive integer has a unique successor, and that every positive integer except the first has a unique predecessor (see, e.g., Enderton, 1977, chap. 4, for a more precise formulation). But how could children grasp these properties by combining the ANS and the object system, neither of which has anything like them? Children also have available at this time the sequence of number words that they have learned to recite—“one,” “two,” “three,” …, up to some largest item, such as “ten.” Adding this sequence into the mix, however, still doesn’t make available the correct integer properties, since the largest numeral on the number-word list has no successor for the children (see Leslie, Gelman, & Gallistel, 2008; Rips et al., 2006; Rips, Bloomfield, & Asmuth, 2008, for more on this theme).

---

7 A similar caution applies to the use of “sets” in the passage quoted earlier. What Spelke (2000) intends is presumably perceptual groupings of objects, not set theory’s formal sets.
The ANS provides systematically misleading information about cardinality, since it associates each new number word with a successively smaller increment in approximate magnitude. The object system yields no information for numerals above “three.” So how do the children learn that a phrase like “four balloons,” for example, refers to exactly four, and “five hippos” to exactly five? As in the case of the mass/count distinction, we’re left wondering where the missing pieces of adult knowledge come from, if not from these building blocks.

4. Core Cognition and Causality

In some well-known research in the 1930’s and 1940’s, the Belgian psychologist Albert Michotte (1946/1963) documented a phenomenon called the “launching effect.” In these experiments, participants viewed an animated 2D scene in which a square moved horizontally until it met a second square. If the second square moved off at the time of contact, observers described the event as the first square “pushing” the second or “causing it to move” or similar causal expressions. But if the demo introduced a spatial gap or a temporal delay between the point at which the first square stopped and the second started, then the observers no longer used causal language but instead described two independent movements. On the basis of these experiments, Michotte concluded that people directly perceive causality in the launching event: They don’t merely see the movements of the objects and then make an inference that a causal interaction must have occurred. They immediately see causality as such (see Rips, 2011, for a review of research and debate stemming from Michotte’s studies).

Michotte’s launching phenomenon lends itself to the idea that people have a launching detector as part of core cognition and that the launching detector supports the adult concept of causality. Although perception of causality could perhaps constitute a core domain of its own, theories of core cognition treat detection of launching in Michotte-style situations as the work of the object system, described in Sections 2 and 3 (see, e.g., Spelke & Kinzler, 2007).

As in our earlier examples, the idea that causality is part of core cognition is backed by experiments on infants. For example, Leslie and Keeble (1987) habituated one group of six-month

---

8 For an online demonstration of the launching effect, see, e.g., cogweb.ucla.edu/Discourse/Narrative/michotte-demo.swf.
olds to a standard launching event involving two rectangles. The infants were then tested on a single trial in which the same event was shown in reverse. That is, if the rectangles moved from left to right in the habituation sequence, they moved from right to left in the test sequence but maintained the same timing. (A film strip was used in habituation, and it was played in reverse at test.) A second group of babies were habituated to a similar sequence, but with a temporal delay inserted between the movements of the two rectangles in order to eliminate the impression of causality. This group was tested on the reversal of this temporally delayed sequence. The logic of the experiment is that reversing both sequences reverses some of their spatio-temporal properties (e.g., the rectangles change direction from left-to-right to right-to-left), but reversing the launching sequence also reverses some of its causal properties: The agent rectangle in the habituation sequence becomes the patient at test, and the patient becomes the agent. So if the infants who were habituated to the launching event are able to grasp the sequence as causal, then they should perceive more properties changing between habituation and test (both spatio-temporal and causal properties change) than for the infants habituated to the non-causal delayed sequence (only spatio-temporal properties change). More change should prompt longer looking times, and that’s what Leslie and Keeble report.

It is possible to debate whether Michotte and his followers are right that we have a hard-wired perceptual device that detects causal events as such, without the aid of inferences in central cognition (see Rips, 2011, for discussion of this worry). But sticking with this article’s plan of not questioning the existence of core systems, I’m going to grant that we have detectors for these events. The question that I do want to raise is what these detectors are detecting. Clearly, these systems aren’t recognizing all and only the causal events that we can think about. They can’t recognize all causal events, because some of them are too small or too far away to be perceived. Even many causal events that we can perceive, we probably don’t directly perceive as causal. For example, I doubt that people directly perceive as causal a train of gears turning or a flame causing a metal bar to glow, despite the fact that they would identify such events as causal on non-perceptual grounds. The same goes for cases of causation by omission (e.g., Fred’s failure to clear the ice from his sidewalk causing Martha to slip). And these core systems can’t be detecting only causal interactions, either. All the classic experiments on launching, including Michotte’s and Leslie and Keeble’s, are non-causal interactions: One
geometric shape doesn’t really cause another to move. Observers are merely looking at simple 2D animated graphics. It’s apparently easy to fool the cause detector.

So what is the core detector detecting? Well, launchings. In addition to launchings, there’s evidence for perceptual detection of a few more kinds of events. One type is what Michotte called “entrainings,” which are like launchings, except that the first object remains in contact with the second. Other perceptual psychologists have proposed pullings, smashings, and burstings as directly perceived (White & Milne, 1997, 1999). It’s possible that we have several cause detectors, one for each of these types of interaction. But in any case, we’re far from having a complete list of everyday causal and perceivable events. No gears turning, birds flying, flames causing metal to glow, burners causing ice cubes to melt, and so on (see Forbus, 2016, for a discussion of possible representations for such events). Even infants appear to have a grasp of the causality of events in which a hidden agent propels a visible patient (e.g., a ball hurtling over a fence) but in which no contact between agent and patient is observable (e.g., Saxe, Tzelnic, & Carey, 2007). Carey (2009, p. 240) suggests in a review of such studies that “not all of an infant’s earliest causal representations are modular. Thus, these results weigh clearly against Michotte’s contention that the perceptual module is the single original source of all true causal concepts….”

To capture a wider range of causal notions, proponents of core cognition have at least two options. One possibility is to enrich the cause detector so that it can do more than simply analyze objects’ intersecting motion paths. For example, Leslie (1994) thinks a core system exists for objects and other bodies (a theory of body) that can attribute force and transmission of force to the objects it’s witnessing. This force property goes beyond what’s available in a purely Michottean module, which yields only “a disembodied or purely visual cause and effect” (Leslie, 1994, p. 128). This enriched system attributes force to animate objects, marking them as dispositional force imparters and likely mechanical agents of an interaction. This system would help explain why nine-month-olds are more surprised to discover an inanimate block or toy train as the source of a launching than to discover a

---

9 Leslie’s (1994) notion of “force” is that of Talmy (1988), rather than actual physical force. According to Talmy (1988, p. 54), “As language treats the concept, an entity is taken to exert a force by virtue of having an intrinsic tendency toward manifesting it—the force may be constant or temporary, but it is in any case not extrinsic.”
hand as the source (Saxe et al., 2007). But this souped-up core component would still be incapable of
dealing with events like fire causing metal to glow or failure to clear away ice causing someone to
slip. So even if proponents of core cognition add force transmission to the cause detector, they won’t
be able to account for many perceptible events that adults regard as causal.

The second tactic that proponents of core cognition could use to explain the wider class of
causal events is to suppose that two or more core components pool their information in central
cognition in order to capture non-contact causality. This is the same strategy we met in examining
Spelke’s theory of the integers. Most core-cognition approaches assign notions of objects and of
agency to separate core components (e.g., Carey, 2009; Spelke & Kinzler, 2007). Even in Leslie’s
(1994) framework (where mechanical agents are part of the theory of bodies, as we’ve just noticed),
intentional agents inhabit a separate theory-of-mind mechanism. People presumably combine
information from these two components in understanding causal intentional events, such as Fred’s
walking to the grocery to buy some milk. But combining information from the object and agent
systems still gives us no way to account for non-contact (and non-agentive) causality. Even simple
and perceivable causal events like ice melting and fire causing metal to glow will be beyond the ken
of the combined systems. And this once again raises the question of how we gain access to knowledge
of causal events of this wider variety. The theory owes us an account of how combining information
from core systems produces correct predictions concerning the events people judge as causal. Why
does combining this information produce adult causal concepts rather than an incoherent clash of
intuitions?

5. The Aftermath

In all the cases we’ve been looking at, the proponents of core cognition assume that infants
come equipped with powerful devices that give them access to high-level information. The
experiments that back core cognition are often persuasive, and the core-cognition theory may be
successful in explaining infant cognition. But the representations from core cognition are also
supposed to be within hailing distance of the relevant adult concepts in these domains. Spelke objects
are supposed to be similar to ordinary physical objects and to drive linguistic distinctions like
mass/count. Approximate number is supposed to be similar to ideas of the natural numbers. And representations of launching are supposed to be similar to representations of general causal events. The idea is that these representations are the building blocks of adult cognition. They’re explanations of how it’s possible for us to acquire the complicated knowledge that we have about these domains.

The cognitive building-blocks idea is an attractive one because it could provide a straightforward explanation of adult thought. Just as compositional semantics offers a principled way to explain the meaning of an infinite number of sentences from the meanings of their words (and from the grammatical rules that combine them), the building-blocks approach may be able to explain the full range of cognitive abilities from the core systems’ basic representations (and to-be-specified combination rules). The theory is supposed to deliver, not a rational reconstruction of these abilities, but a psychologically real depiction of where these abilities come from. If we knew the contents of the core systems and the way people fit these contents together, then—provided these contents and combination principles are complete—we’d have a full story about the nature of cognition.

But the problem is that the core systems’ representations are far from complete. Spelke objects exclude too many things that are obviously part of adults’ ideas of physical objects and that we name with count nouns. The land of Spelke objects has no mountains, houses, trees, lakes, or roads. The arithmetic of the core number systems has no 1,003 (at least, not in a way that we could distinguish from 1,004). The causality of Michottean cause detectors has no non-contact causality. Of course, we don’t necessarily expect babies to have sophisticated concepts in these domains. So we could live with this state of affairs if we knew a way to bridge between core cognition and adult knowledge. But we don’t.

I think it’s also clear in retrospect what the source of this problem is. Core cognition comprises encapsulated systems (see Section 1) that limit the information these systems can obtain. The input to a core system is sensory/perceptual information, and although it can perform internal inferences based on this information, it has no access to either central beliefs or other core representations. This self-containment is what allows infant researchers and perceptual psychologists to study these systems. But the knowledge we have of these domains isn’t limited in the same way. Our concepts of individual objects allow us to trace them over the course of decades and to refer to
them with count nouns, even when we have little perceptual contact with these items (e.g., Rips, Blok, & Newman, 2006). Adult concepts of numbers and causality both depend on information to which core systems have no access, assuming current descriptions of these systems. The story about how people learn more advanced concepts is supposed to be that they get them by combining core cognition via language. That’s the essence of the building-blocks idea. But combining two or more core representations doesn’t by itself produce the abstract representations that underlie adult thinking. We’ve seen an example of this failure in Section 3: Combining object representations with approximate-number representations doesn’t produce veridical representations of the integers.

It’s tempting to put this point by saying that core representations are perceptual whereas adults’ concepts of objects, numbers, and causality are not. Fodor (1984) thinks that the boundary between perceptual modules and central cognition marks the distinction between what people can observe and what they can only infer, and the distinction between Fodorian modules and central cognition may be the best way we have to mark the boundary between perception and (higher) thought. If core systems are “cognitive modules in Fodor’s sense” (see the quotation from Spelke in Section 1), then there’s some reason to regard core systems as perceptual systems. However, Spelke (1988) denies that core systems are perceptual (see also Carey, 2009) and, in particular, that the core object system is a perceptual one. According Spelke’s account, perception yields only continuous surface properties of the spatial array (and their movement) but does not divide them into individual objects: “Perceptual systems do not package the world into units. The organization of the perceived world into units may be a central task of human systems of thought… The parsing of the world into things may point to the essence of thought and to its essential distinction from perception” (Spelke,
Because the job of the core object system is “parsing the world into things,” core object cognition can’t be perceptual, according to Spelke’s theory.\(^{10}\)

But the difficulties with core cognition don’t depend on whether we should regard these systems as perceptual. Instead, the deficiencies have to do with restrictions on the systems’ representational capacities. To make it plausible that infants and some nonhuman animals have these core systems, proponents of core cognition have to place strict limits on how smart these systems can be. And because these systems are encapsulated, the systems can’t learn to overcome these limits through corrective feedback from central beliefs. That’s why children’s breakthroughs in understanding have to come through activities in central cognition by means of conceptual combination. But if core representations provide the only input to the combinatorial building-block process, then they need to be complete—sufficient to account for all adult concepts.

No matter how fuzzy the boundary between perception and cognition, the core systems’ representations must be relatively less complex than those of adult cognition. Otherwise, proponents of core cognition would have trouble accounting for obvious limits on infants’ cognitive abilities. Since each core system has limitations of this same sort, combining the output from these systems doesn’t necessarily provide the representational power that adult knowledge exhibits. Core cognition’s building-block view of development would be reasonable if it had all the pieces it needed to assemble adult concepts. But pieces seem to be missing in the domains we’ve considered here, and the theory has no obvious suppliers for these parts.

Where do the missing pieces come from, if not from core systems? As mentioned earlier, Carey’s (2009) radical conceptual change is one possibility (but for doubts, see, e.g., Fodor, 2010;)

\(^{10}\) It might be worth noting that the distinction that Spelke (1988) is drawing in the quoted passage seems closer to what many psychologists would identify as the difference between sensation (which is largely non-inferential) and perception (which is inferential), rather than between perception and thought. (See, e.g., Yantis, 2014, p. 2: “Sensation: The initial steps in the perceptual process, whereby physical features of the environment are converted into electrochemical signals…Perception: The later steps in the perceptual process, whereby the initial sensory signals are used to represent objects and events so they can be identified, stored in memory, and used in thought and action.”) Similarly, most contemporary perceptual psychologists would probably dispute the claim that we can only perceive continuous surface layouts and not individual objects (as the quotation from Yantis attests). But this isn’t the place to pursue these issues. For one thing, the present point (as discussed in the next paragraphs) doesn’t depend on them. For another, Spelke’s (1988) paper preceded her development of the core cognition theory, and it is not clear whether the opinions she expresses there represent her final thoughts on the distinction between perception and conception.
Rey, 2014; Rips et al., 2006). Other potential sources are concepts from central cognition, either learned or, more likely, innate since it’s hard to come up with a convincing theory of how they could be learned. In some of the domains we’ve considered, it’s also fairly clear what the missing information is. For example, in the case of knowledge of the positive integers, it’s the idea of a successor function (see Leslie et al, 2008; Rips, Bloomfield, & Asmuth, 2008). Supposing that the missing conceptual primitives come from outside core cognition, then we have the ingredients we need for remedying core cognition’s incompleteness.

The main question about innate, non-core conceptual primitives is why they don’t come online early, around the time that core knowledge itself does. For example, why do kids take such a protracted period of time—a year or more—to understand the quantifier meanings of the first few numerals for the integers? One answer might be competition among rival ways of using these primitives. Children have to learn during this period that distinct structures are appropriate for different domains: finite linear lists (e.g., for the alphabet), circular lists (e.g., for the months of the year or days of the week), branching structures (e.g., for taxonomic or meronomic systems), among others. So, although kids may be capable of learning the positive integers at an early point (in the sense of having the right primitives to construct them), they may have to wait for appropriate evidence to figure out which structure is correct. Assuming conceptual primitives outside core systems isn’t the only way to help core cognition out of its difficulties, but it may be the simplest way.


Spelke, E. S. (2011a). Quinian bootstrapping or Fodorian combination? Core and constructed knowledge of number. *Behavioral and Brain Sciences, 34*, 149-150.


Table 1

*Some Examples of the Mapping between the Solid/Nonsolid Distinction among Entities and the Mass/Count Distinction among Nouns in English*

<table>
<thead>
<tr>
<th>Categories of Core Cognition</th>
<th>Grammatical Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Solids:</strong></td>
<td></td>
</tr>
<tr>
<td>cat</td>
<td>“cat”</td>
</tr>
<tr>
<td>toaster</td>
<td>“toaster”</td>
</tr>
<tr>
<td>gold</td>
<td>“soufflé”</td>
</tr>
<tr>
<td>cows/cattle</td>
<td>“cow”</td>
</tr>
<tr>
<td>jewelry</td>
<td>“suggestion”</td>
</tr>
<tr>
<td>furniture</td>
<td></td>
</tr>
<tr>
<td><strong>Non-solids:</strong></td>
<td></td>
</tr>
<tr>
<td>mud</td>
<td>“mud”</td>
</tr>
<tr>
<td>water</td>
<td>“water”</td>
</tr>
<tr>
<td>soufflé</td>
<td>“gold”</td>
</tr>
<tr>
<td></td>
<td>“cattle”</td>
</tr>
<tr>
<td></td>
<td>“jewelry”</td>
</tr>
<tr>
<td></td>
<td>“furniture”</td>
</tr>
<tr>
<td></td>
<td>“advice”</td>
</tr>
</tbody>
</table>
Table 2

Some Examples of the Mapping between the Spelke-object/Nonobject Distinction among Entities and the Mass/Count Distinction among Nouns in English

<table>
<thead>
<tr>
<th>Categories of Core Cognition</th>
<th>Grammatical Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spelke objects:</td>
<td>Count nouns:</td>
</tr>
<tr>
<td>cat</td>
<td>“cat”</td>
</tr>
<tr>
<td>toaster</td>
<td>“toaster”</td>
</tr>
<tr>
<td>soufflé (?)</td>
<td>“soufflé”</td>
</tr>
<tr>
<td>cows/cattle</td>
<td>“cow”</td>
</tr>
<tr>
<td>jewelry</td>
<td>“suggestion”</td>
</tr>
<tr>
<td>furniture</td>
<td>“tree”</td>
</tr>
<tr>
<td></td>
<td>“house”</td>
</tr>
<tr>
<td></td>
<td>“ceiling”</td>
</tr>
<tr>
<td></td>
<td>“road”</td>
</tr>
<tr>
<td></td>
<td>“lake”</td>
</tr>
<tr>
<td></td>
<td>“nose”</td>
</tr>
<tr>
<td>Non-(Spelke) objects:</td>
<td>Mass nouns:</td>
</tr>
<tr>
<td>mud</td>
<td>“mud”</td>
</tr>
<tr>
<td>water</td>
<td>“water”</td>
</tr>
<tr>
<td>gold (?)</td>
<td>“gold”</td>
</tr>
<tr>
<td>tree</td>
<td>“cattle”</td>
</tr>
<tr>
<td>house</td>
<td>“jewelry”</td>
</tr>
<tr>
<td>ceiling</td>
<td>“furniture”</td>
</tr>
<tr>
<td>road</td>
<td>“advice”</td>
</tr>
<tr>
<td>lake</td>
<td></td>
</tr>
<tr>
<td>nose</td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Habituation trials and test trials from Hespos et al. (2009, Experiment 1).
**Figure 2.** Mean looking time to test stimuli for infants habituated to a liquid and infants habituated to a solid. Error bars represent one standard error of the mean. Data from Hespos et al. (2009).
Figure 3. Habituation and test displays from Xu and Spelke (2000).