Chapter 3. Core knowledge of objects

3.1. Core knowledge

Chapter 2 challenged a widely shared assumption about the ontogenetic origins of human conceptual understanding, the assumption that the initial stock of representations are limited to perceptual or sensori-motor primitives. My characterization of an alternative to the empiricist picture draws especially on the writings of Renee Baillargeon, Randy Gallistel, Rochel Gelman, Alan Leslie, and Elizabeth Spelke. These writers and I believe that human cognition, like that of all animals, begins with highly structured innate mechanisms designed to build representations with specific content. Following Spelke et al. (1992) I call these real world content domains "core domains," and the mental structures that represent them "core knowledge."

Core knowledge has several properties. First, core knowledge has conceptual content; it cannot be characterized in terms perceptual or sensorimotor primitives. Also, as we will see here, core knowledge is conceptual in a second sense; it includes representations that are explicit in the sense of being accessible, attended, and available to thought and to guide action. Second, core knowledge is articulated in terms of representations that are created by specialized perceptual input analyzers. As in the empiricist view of concepts, part of what determines the content of a concept in core knowledge is a causal connection between the real world entities in its domain and the concept. The causal connection between a concept in core knowledge and its referents is mediated by the operation of perceptual input analyzers that have been constructed, through natural selection, specifically for the purpose of representing certain classes of entities in the world. Third, the perceptual analysis devices that identify the entities that fall under core domains continue to operate throughout life. Core knowledge is elaborated during development, as core knowledge systems are learning devices, but it is never rendered irrelevant. It is never overturned nor lost as are earlier intuitive theories that are replaced by later, incommensurable ones. Fourth, some core knowledge (including that of objects) is shared by other animals. At least some early developing cognitive systems in humans have a long evolutionary history. Fifth, core knowledge acquisition is supported by domain specific learning devices. In most of these respects, core knowledge representations resemble perceptual representations and differ from those in almost all other domains of human conceptual knowledge, including intuitive theories.

In this chapter, I continue with the example of object representations. It is an empirical claim that there are systems of core knowledge. That is, it is an empirical claim that some systems of knowledge representations have the properties outlined above. Object representations have all these properties, and in laying out this case I both illustrate the characteristics of core knowledge as well as the evidence for it. I begin with one of the properties not touched on in Chapter 2: the representations that articulate core knowledge of objects continue to operate throughout life. This feature is central to the core knowledge thesis, for it is one of the respects in which core knowledge differs from theoretical knowledge. Theories change, sometimes radically, such that even basic ontological commitments are revised. Indeed, theories may and do overturn, at an explicit level, tenets of core knowledge, even while core knowledge representations are also still computed. Surely theories of objects change; the discovery that objects are made up of particulate matter, for instance, is not part of core knowledge and even violates the solidity constraint. So what is the evidence that core knowledge of objects persists throughout the life span?

3.2. Two (apparently independent) research programs for the study of object representations.

As Chapter 2 illustrates, the study of the *development* of object representations has had a long intellectual history, as it is a case study that has animated arguments between empiricists and rationalists as long as this debate has raged. The study of the *adult* perceptual mechanisms that create object representations also has had a long intellectual history, at least back to the Gestalt psychologists, but continuing today in the study of what is called "mid-level object based attention" (see Scholl, 2001). The two research communities (infant cognition and adult visual cognition) have been largely separate, motivated by slightly different theoretical concerns. The thesis that core knowledge of objects continues to operate throughout the life span is supported by the discovery that the representations these two different research communities have been studying are one and the same.

Consider for a moment the problems that drive the work on adult object perception. Sensory input is continuous. The array of light on the retina, even processed up to the level of Marr's 2 1/2 D sketch (Marr, 1982), is not segregated into individual objects. Yet distinct individuals are provided by visual cognition as input into many other perceptual and cognitive processes. It is individuals we categorize into kinds; it is individuals we reach for; it is individuals we enumerate; it is individuals among which we represent spatial relations such as "behind" and "inside"; and it is individuals that enter into our representations of causal interactions and events. Because of the psychological importance of object individuation, the problem of how the visual system establishes representations of individuals from the continuous input it receives has engaged psychologists for almost a century.

The two literatures, that on mid-level object-based attention (mid-level because the representations fall between low level sensory processing and high level placement into kind categories) and that on object representations in infancy, involve parallel problems, including uncovering the bases of object individuation and numerical identity. Recently, many have suggested that both communities have actually been studying the same psychological mechanisms; that is, that the object representations of young infants are identical to those that are served up by mid-level object-based attentional mechanisms (Carey and Xu, 2001; Leslie, Xu, Tremoulet, & Scholl, 1998; Scholl and Leslie, 1999; Simon, 1997; Uller, Carey, Huntley-Fenner, & Klatt, 1999). I endorse this proposal, with an important emphasis on *young*. If the proposal is correct, then this single system of object representation exemplifies three properties of core knowledge: 1) mid-level object representations are created by encapsulated, dedicated input analyzers, 2) they are innate (or at least very early developing) and 3) they continue to articulate our representations of the world throughout the life span.

3.3 How adult object individuation works

The story is complicated by the existence in adults of at least two distinct representational systems underlie object individuation. The first is the mid-level visual system that assigns spatiotemporal indices to attended objects and creates what are called "object-files," symbols for attended objects that provide a spatiotemporal address for the object in the world. Representations of the properties of an attended objects can be bound to the object files that represents it, and these can be updated as the object changes through time. Attention is required for binding properties in object files. It is this system that is identified with object representations in young infants. The object-file system privileges spatiotemporal information in the service of individuation and establishing numerical identity. Individual objects are coherent, spatially separate and independently movable, spatiotemporally continuous entities.[1] Other perceptual information is drawn upon in the service of edge and surface assignment (as in Gestalt principles of figure/ground segregation) but only secondarily in decisions about numerical identity, used only when spatiotemporal evidence is ambiguous or absent. For example, if we see two objects, a cat and dog, emerge together from behind a wall and return, we have spatiotemporal evidence that there are two. If they reemerge, we again have evidence for two spatially distinct and separately moveable entities. But we must draw on kind and property information to determine which, if either, of the second set is the same one as the original dog and which is the same one as the original cat. If we had not lost sight of the original two, we could and would trace numerical identity by keeping track of the continuous trajectory. This information trumps all other; if we saw the cat transforming into a dog-shaped entity, we would establish a representation of a single individual (neither a real cat nor a real dog, our conceptual system would tell us) that began with cat-like properties and ended with dog-like properties.

A second system available to adults differs from mid-level object individuation and tracking in that it draws on kind information for decisions about individuation and numerical identity. For adults, individuation might be based on kind information when no relevant spatiotemporal evidence is available, as in the above example. Another example: we decide that the cup on the window sill is the same one we left there yesterday, but the cat on the window sill is not the same individual as the cup we left there yesterday. In the adult conceptual system, kind information often overrides spatiotemporal continuity, as when we decide that a *person* ceases to exist when she dies, in spite of the spatiotemporal continuity of her *body*. Evidence concerning the perceptual properties of entities is relevant to individuation at the conceptual level, but identity/difference in perceptual features are not sufficient to drive judgments about numerical identity. Our inferences concerning the relevance of perceptual properties to individuation decisions are kind-relative; a 6" long puppy may be the same individual as a 4' long dog a year later, but a tiny fountain pen will not be the same individual as a large one a month later. Color differences do not signal distinct individual chameleons, but barring dyes, they do ordinarily signal distinct individual cats.

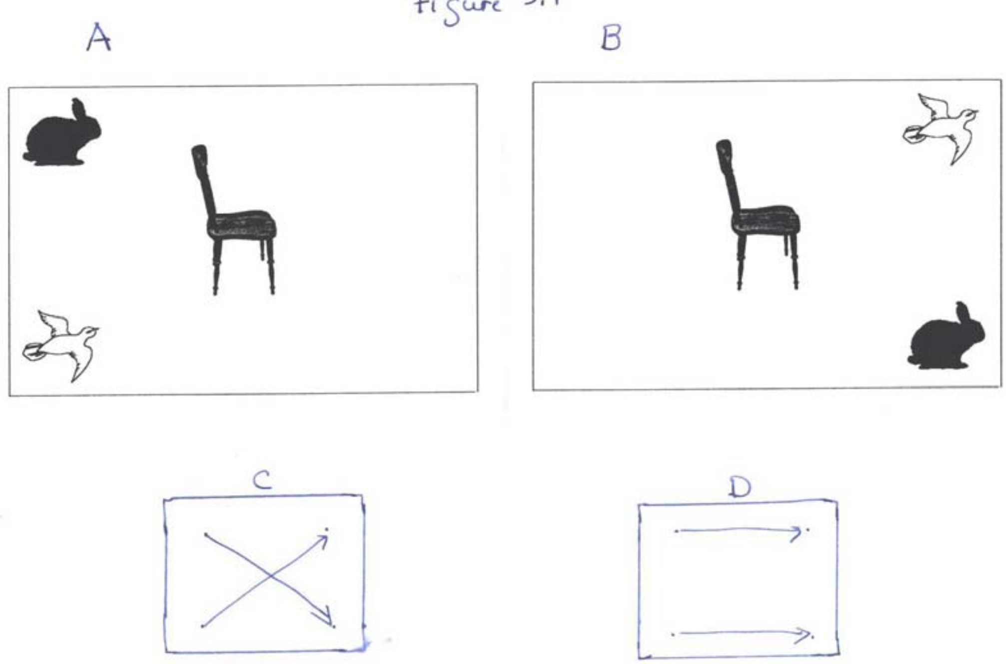
Figure 3.1 illustrates the operation of the two systems in establishing numerical identity. Imagine that you lose perceptual contact with the scene in Panel A, and return 5 minutes later to view the scene depicted in Panel B. How would you describe what has happened? I assume you would say that the rabbit has moved from above and to the left of the chair to below and to the right of it, while the bird has moved from the bottom left

---

[1] The exact characterization of the individuals that are indexed in object tracking experiments and represented by object-files is a matter still up for grabs. See Scholl, Pylyshyn & Feldman (2001) and Scholl et al., 2003 for a investigagtions into what individuals can be indexed and tracked in multiple-object-tracking (MOT) studies. It seems likely that groups of spatially separate entities undergoing common motion are construed as individuals in these studies, and non-cohesive or radically flexible individuals are not.

to the top right. That is, you would report the movements of the individuals as in Panel C. In this account, numerical identity is being carried by kind membership; it is the rabbit and the bird each of whom you assume has moved through time. The conceptual, kind-based, system of individuation is responsible for establishing the object tokens in this case. Now imagine that the chair is replaced by a fixation point, and Panels A and B are projected one after the other onto a screen, while you maintain fixation on the common fixation point. If the timing of the stimuli supports apparent motion, which individuals do you see in motion? Rather than seeing a bird and a rabbit each moving diagonally, you see two individuals each changing back and forth between a white bird-shaped object and a black rabbit-shaped object as they move side to side, as in Panel D. . The visual system that computes numerical identity of the objects that undergo apparent motion in cases such as this works to minimize the total amount of movement; this system takes into account property or kind information only when spatio-temporal considerations are equated (see Nakayama, He, & Shimojo, 1995, for a review).[2] The mid-level object tracking system is responsible for establishing the object tokens in the case of apparent motion, and it settles on a different solution than does the kind-based system.

Figure 3.1



Chapter 6 concerns, in part, the ways in which studies on object individuation in infancy support the architectural distinction between kind-based object individuation and

[2] Suppose there are two possible solutions to establishing numerical identity among the individuals in two sets of two, as in Figure 3.1. If the total motion involved in the two solutions is the same, then property identity tips the balance toward one solution over the other.

mid-level object file based individuation. Here I focus on arguments that the mid-level object-file system underlies object representations of very young infants. As argued by Scholl and Leslie (1999) and by Carey and Xu (2001), the conjecture that object representations in young infants are in fact the object files of adult object-based attention rests on three distinct considerations. First, and most importantly, both systems of representation privilege spatiotemporal information in computations of numerical identity. Second, both systems of representations are subject to the same set size limitations of number of objects that may be simultaneously attended and represented (on the order of 3). Third, the object representations of both systems survive occlusion, and object tracking is sensitive to the distinction between loss of visual contact that signals cessation of existence and loss of visual contact that does not.

3.4 <u>Signal properties of the mid-level object tracking system</u>

To fully review the literature on object-based attention is beyond the scope of this chapter; luckily, there is a recent special issue of the journal *Cognition* (Scholl, 2001) that does the job. A variety of experimental paradigms support the following generalizations: 1) although attention may be directed to locations, it is often the case that attention is allocated to individual objects that are traced through time and space. 2) Attended objects in the world are assigned spatial indexes and traced through time on the basis of spatio-temporal continuity. 3) Attention is required to bind features in object-files, and represented features may be updated as objects change over in time. 4) Only 3 indices may be assigned at any one time; that is, only 3 or 4 individual objects may be attended to at any one time.

Consider the first generalization, that attention is allocated to objects rather than locations. In the paper in which they coined the term “object-file,” Kahneman, Triesman and Gibbs (1992) exposed participants in their studies to letters in boxes, after which the boxes moved to new locations. See Figure 3.2. The participants’ task was to then read the letter in a cued box. What primed (or interfered with) letter identification was whether the letter in the spatio-temporally defined *object* was the same or different from before, not whether the letter that had been in the *same location in space* as the currently cued letter was the same or different from before. Apparently the participants established representations of individual objects and bound features to them (in this case the identity of the letter contained within), the bound features remaining as they traced the identity of the objects through time on the basis of their spatiotemporally continuous motion.

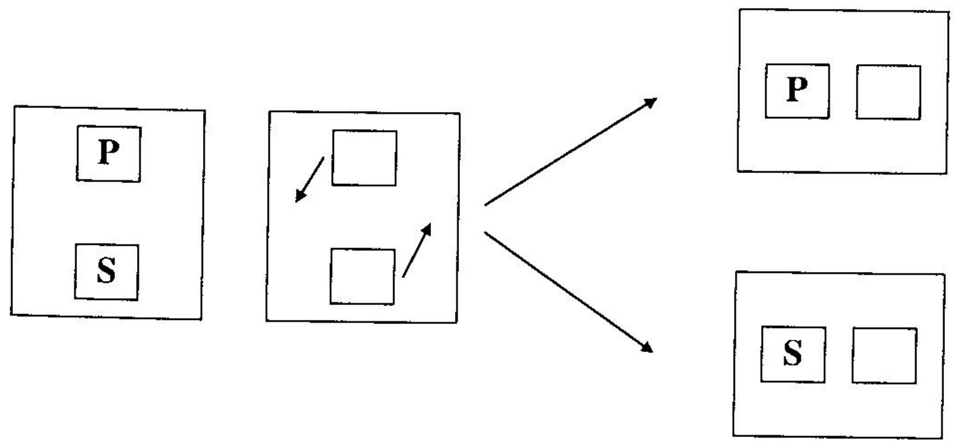


Figure 3.2

Pylyshyn and his colleagues (Pylyshyn, 2001; Pylyshyn and Storm, 1988) developed a fruitful paradigm, multiple object tracking (MOT), to explore the processes that index attended objects. In these studies, participants are shown a display as in Figure 3.3 consisting of many individual figures (e.g. 8 in this example). A subset of these figures are highlighted, indicating the set that the participant is to track, and then those again become indistinguishable, featurally, from the rest. The entire array is then put into motion; the objects move randomly and independently from each other, and the observers task is to keep track of the attended set. After a period of tracking, the motion is stopped and the observer must indicate whether a highlighted individual is part of the attended set.[3]

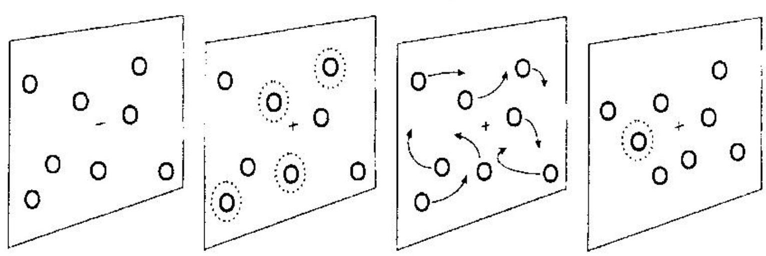
There are several notable results from MOT studies that provide empirical support for the generalizations listed above about mid-level object tracking. First, the number of objects that may be tracked in parallel is sharply limited. Performance is excellent when sets of one, two, three are tracked, but falls apart thereafter.[4] This is part of the empirical basis for the claim that there is a limit on the number of indexes that may be assigned at any one time. Second, consistent with the claim that object tracking is based on spatiotemporal continuity, and that feature changes do not signal the opening of new object files, object tracking in the multiple object tracking (MOT) studies is not disrupted by indexed objects changing color, size, shape or kind during their motion (Pylyshyn, 2001). Additionally, a recent study by Scholl, Pylyshyn & Franconeri (date) underscores the primacy of spatio-temporal information in the establishing and tracking of object files. In the MOT paradigm, if the motion all the objects is stopped, at which time one of the tracked objects disappears, the participants can indicate its last seen location and direction of motion. But if objects are changing properties during tracking, participants are not aware of the last seen color or shape of a tracked object. Not only is

[3] In other versions of the task, no individual is highlighted after the motion stops, but rather participants must indicate the whole set. The two versions yield comparable results.
[4]Scholl (2001) shows that above chance performance when tracking 4 or 5 objects can be explained on the assumption that subjects are tracking 3 and guessing.

spatiotemporal continuity the basis of tracking, participants have conscious access to the spatiotemporal address of a currently attended object, but not always to other features of the indexed object.

In sum, the computations that maintain indexes to attended objects rely heavily on spatiotemporal information; objects are tracked on the basis of spatiotemporal continuity. Once an object-file is opened, features may be bound to it, and updated as they change through time. (The Scholl et al. study just described shows that features are not automatically bound in open object-files, perhaps because of the high attentional demands of tracking 3 independently moving objects at once). These generalizations hold for the young infant's object representations as well, the point to which we now turn.

Figure 3.3



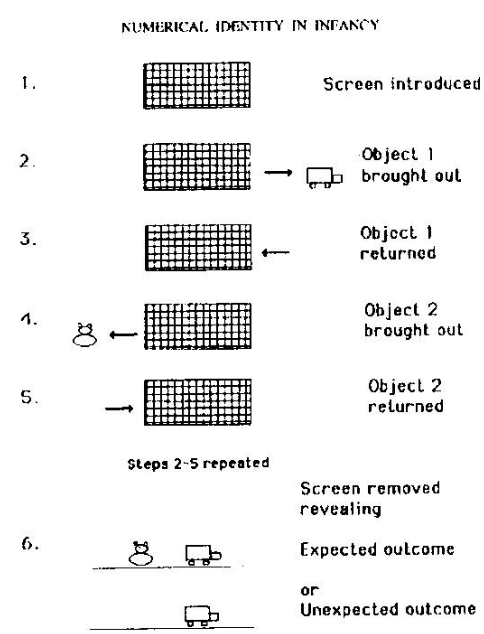
3.5. <u>Infant object individuation: Primacy of spatiotemporal information.</u>

A lovely infant study by Kirkham and Johnson (date) is modeled on the object-file experiment of Kahnemann et al. depicted above (Figure3.2). They exposed 7-month-old infants to a computer display consisting of two identical boxes arranged vertically. While the infant watched, a duck was revealed in one box and quacked, the box then returning to its featureless state. Then a bell was revealed in the other box and rang and this box returned to its featureless state. After seeing these events repeat several times, the identical boxes went slowly into motion until they were arranged horizontally, equally distant from the midpoint of the display.. At this point, either a quack or a ring was played and the dependent measure was where the infants looked. They looked to the box in which the matching object/sound had been revealed during habituation. Analogously to the Kahnemann et al. priming studies, these infants had established representations of individual objects, bound features (the identity of the sounds that emerged from them), maintaining the binding as they traced the identify of the objects through time on the basis of spatiotemporally continuous paths.

Whereas this experiment suggests that infants have object-file representations available to them, the thesis under consideration here is that the object representation experiments described in Chapter 2 reflect exactly these representations. The Spelke et al. (1995) and the Wynn (1992) experiments show that infants as young as 4 months of age draw on spatiotemporal information in object individuation and tracking. Because the objects in those studies were perceptually indistinguishable from each other, evidence of spatio-temporal discontinuity must have driven representations of two distinct objects. However, these studies do not show that spatiotemporal information is privileged, for they did not explore whether infants could also use perceptual property differences (e.g., red vs. blue, cup-shaped vs. duck shaped) or kind distinctions (e.g., cup vs. duck) as a basis for object individuation.

Recent studies suggest that in many conditions in which spatiotemporal information is sufficient for object individuation, young infants fail to use property or kind differences among objects for this purpose (Xu & Carey, 1996; Van de Walle et al., 2000).. Imagine the following scenario: One screen is put on a puppet stage. A toy duck emerges from behind the screen and returns behind it, and then a ball emerges from behind the same screen and then returns (Figure 3.4). How many objects are behind the screen? For adults, the answer is clear: At least two, a duck and a ball. But since there is only a single screen occluding the objects, and because we never see both objects at once, there is no clear spatiotemporal evidence that there are two objects. We must rely on our knowledge about perceptual properties or object kinds to succeed at this task.

Figure 3.4



In our studies, 10- and 12-month-old infants were shown the above event. The objects contrasted in kind and properties (in the above example, a duck vs. a ball; a yellow, duck-shaped, plastic object vs. a red, round, rubber objects). Some objects were toy models (e.g., truck, duck, elephant) where others were from highly familiar everyday kinds (e.g., cup, bottle, book, ball). On the test trials, the screen was removed to reveal either the expected outcome of the two objects or the unexpected outcome of only one of them. If infants have the same expectations as adults, that these kind or property differences signal two distinct objects, they should look longer at the unexpected outcome of a one object. The results, however, were surprising: 10-month-old infants failed to draw the inference that there should be two objects behind the screen, whereas 12-month old infants succeeded in doing so.

Control conditions established that the method was sensitive to infant representations of distinct individuals. Ten-month-old infants succeeded at the task if they were given spatiotemporal evidence that there were two numerically distinct objects, e.g., if they were shown the two objects simultaneously for 2 or 3 seconds at the beginning of the experiment. Furthermore, Xu and Carey (1996) showed that infants are sensitive to perceptual or kind differences under the circumstances of their experimental paradigm: It takes infants longer to habituate to a duck and a car alternately appearing from each side of the screen than to a single car (or duck) repeatedly appearing from behind the screen. Ten-month-old nfants are sensitive to the property or kind differences, but they do not use these differences as a basis of object individuation, under these circumstances.

In this task, 10-month-old infants failed to draw on kind-based individuation over a wide range of kinds, such as *duck, truck, animal, vehicle, cup, bottle,* and *book.*. They also failed to draw on perceptual contrasts, such as the contrast between being yellow, duck-shaped, and rubber vs. being red, car-shaped, and metal. Other than spatiotemporal features of boundedness and continuity of paths, the perceptual properties of objects that infants under 10 months of age are sensitive to may be irrelevant to object individuation under conditions of spatiotemporal ambiguity concerning numerical identity of occluded objects.[5]

Van de Walle, Carey, & Prevor (2000) sought convergent evidence for the claim that infants below 12-months of age are more sensitive to spatiotemporal information than kind information in computations that establish numerical identity among objects. In these studies, a manual search measure was used instead of looking time. Ten- and 12-month-old infants were trained to reach through a spandex slit into a box into which they could not see in order to retrieve objects they had seen removed from and replaced into it. Three types of trials were contrasted: 1) 1-object trials, in which the same object (e.g., a toy telephone) was removed from the box twice, and replaced twice, 2) 2-object trials in which objects of different kinds (e.g., a telephone and a ball) were removed one at a time and replaced in the box, such that the two were never seen together, and 3) 2-object trials in which two different objects were removed one at a time, but shown together before being returned into the box. In the second type of trial, infants must rely on property or kind contrasts as a basis for object individuation; the third type provided spatiotemporal information as well.

After one of these introductions to the contents of the box, the box was pushed into the child's reach, and patterns of search revealed how many objects the child had represented in it. In these experiments, we surreptitiously removed one of the objects of 2-object trials, so there was in fact only one object in the box. We could then measure persistence of search for a second object. The question was whether infants search for a second object after having retrieved one on 2-object trials (types 2 and 3) but not on 1-object trials. Both 10- and 12-month-olds differentiated the 1- and 2-object trials when

---

[5] Other laboratories have replicated these findings (e.g., Wilcox & Baillargeon, 1998a, Bonati and Mehler, 2002) In our original writings on this topic, Xu and I made a blanket claim that infants under 11 or 12 months could never use property or kind information in the service of object individuation. This claim is too strong. Nonetheless, some of the apparently conflicting data have alternative explanations. See Xu & Carey, 2000, for a discussion of some apparently conflicting data from Needham & Baillargeon, 1998; Wilcox, 1999; Wilcox & Baillargeon, 1998a, 1998b). The resolution of these controversies is taken up in Chapter 6. For the point I am making here, all that is necessary is young infants fail to make decisions about numerical identity on the basis of property/kind distinctions under conditions in which they succeed when provided unambiguous spatiotemporal information. This pattern of findings has been widely replicated in the literature.

given *spatiotemporal* evidence for two objects. Twelve-month-olds also succeeded when given *property/kind* information alone. In contrast, the 10-month-old failed in this condition; their pattern of reaching on the 2-object trials was the same as on the 1-object trials. Ten-month-olds failed to use kind differences such as *telephone, duck* or *car, book* or property differences such as *black, yellow, telephone-shaped, duck-shaped, rubber, plastic* to establish representations of two numerically distinct objects in the box.

Thus, results from these manual search studies are completely consistent with those from the looking time studies. Two important conclusions follow from these data. First, they support the identification of the young infants' object representations with those of the mid-level object tracking system, for they show that under these circumstances at least 10-month-old infants fail to draw on property/kind information in the processes that establish whether an attended object is numerically identical or different from another, under conditions in which they do draw on spatiotemporal information for this purpose. Second, they are consistent with the claim that kind-based object individuation is architecturally distinct from mid-level object indexing and tracking (see Figure 3.1 and surrounding discussion). They support the possibility that a second system of object individuation, a kind-based system, emerges at around 12 months of age. Of course, the data presented so far do not show that kind distinctions rather than property distinctions underlie the older infants' success; Chapter 6 takes up this issue and discusses the mechanisms that might underlie the construction of a new representational system, kind-based object representations, that goes beyond the core knowledge system of mid-level object files and object indexing.

3.6. Infant object tracking: Set size limitations

Pylyshyn's MOT paradigm provides direct evidence regarding the number of objects that may be simultaneously indexed and tracked through time. Although various task variables affect the set size at which performance is virtually errorless, a good approximation is that about three objects are the adult limit. Results from several paradigms suggest that the limits on young infants' parallel individuation of objects is in the same range. In the interest of space, we mention just two lines of relevant work.

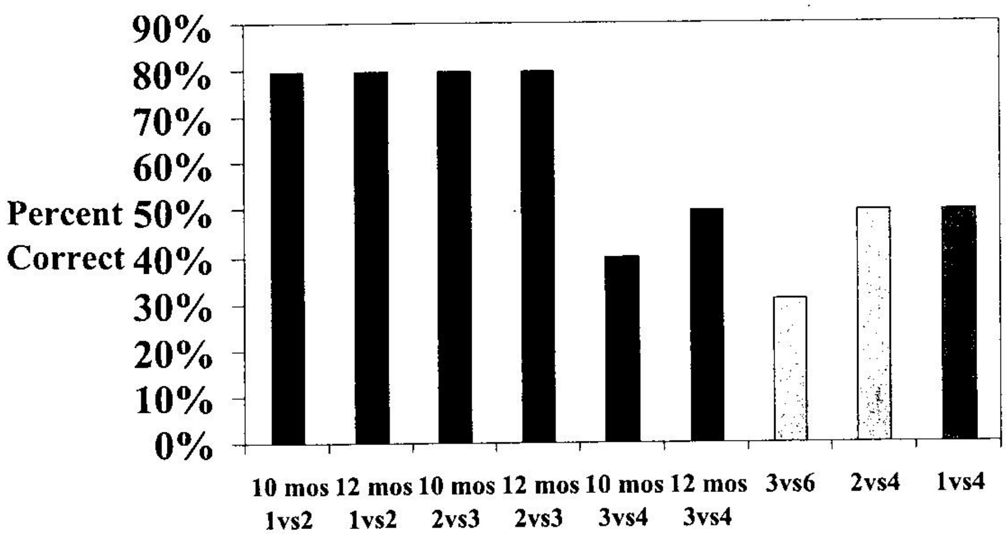
Consider the manual search paradigm described above, in which infants search for objects in a box into which they can reach but cannot see. Above we were concerned with the criteria of individuation infants use in order to establish representations of the distinct individuals in the box. But the method can also be used to explore whether there is an upper limit on the number of objects infants can represent in the box. In these experiments, only spatiotemporal evidence for individuation is provided; the objects are identical to each other. Infants watch, for example, as either one or two objects are placed into the box, the box is given to them and the second object on 2-object trials is surreptitiously removed, such that there is always just one object in the box. Success consists of persistent search for a second object on 2-object trials (in which a second object is expected) in the face of no reaching or merely cursory reaching on 1-object trials (in which the box is expected to be empty, because the only object seen to be put in the box has been retrieved.) Several studies have found success with 1 vs. 2 trials , at both 12 and 14 months of age (Van de Walle et al., 2000; Feigenson and Carey, 2003; Feigenson and Carey, under review). In two series of studies, Feigenson and I have explored the upper limit on infants' representations under these circumstances. We have found that infants succeed just in case the total number of objects in the box is three or less, but performance falls apart at four. For instance, infants of these ages search for more objects if they have seen three go in and have retrieved two, if they have seen three go in and have retrieved one, but they fail if they have seen four go in and have retrieved two or even if they see four go in and have only retrieved one. Apparently, infants cannot form a representation of a set of four individuals under these circumstances.

Stop and think about these data. They tell us something important about how number is represented under these circumstances, lessons I shall return to when I discuss number representations in Chapters 5 and 7. Under these circumstances, infants apparently are not forming a summary representation of the cardinality of the set, even approximately, because they cannot distinguish four objects from one object. It isn't that they fail to represent anything in the box when there are four objects, for they reach in and retrieve one or two. If you give a child a box into which there have been four objects placed, and you have removed all four of them, children search more persistently than if you hand them a box known to be empty. They distinguish 4 from 0, but not from 1 or 2. This is a puzzling result. Even if infants are creating object-file representations in short term memory, and are limited to three, why don't they represent three of the objects in the box and thus differentiate the 4-object trials from the 1- or 2-object trials? One way to think about this result is that objects must be stably indexed in order for object-files to be formed and the limit of three indices means that infants scan the four objects without indexing any three of them. Because the number of objects presented in four object trials exceeds the number of indices, no object files are created. Some other system must represent "something" in the box. In the MOT paradigm, when tracking too many objects so that we lose track of all of them, we still remember that we were tracking objects.

Studies of spontaneous choice between two sets of objects provide data entirely convergent with those from the manual search paradigm (Feigenson, Carey, and Hauser, 2002; Feigenson and Carey, under review). Ten- and 12-month-old infants were shown a certain number of graham crackers placed into one bucket, one at a time, and a different number placed into another bucket, also one at a time. The infants could not see the crackers in the buckets. After watching the crackers being placed, the infants were allowed to crawl to one or the other. At issue was whether they would go to the bucket with the larger number of crackers. This is what they did, when the choice is 1 vs. 2 or 2 vs. 3. Performance was at chance at 3 vs. 4, 2 vs. 4, 3 vs. 6, and even 1 vs. 4 (Figure 3.5). Performance fell apart if one of the sets exceeded 3. Just as in the above experiment, it isn't that the infants represented nothing when there were 4 or more objects—performance was random, not systematically in favor of the smaller number in 1 vs. 4, 2 vs. 4, and 3 vs. 6. Apparently, when there were 4 or more graham crackers in a bucket, they represented "graham cracker in that bucket" but failed to establish a representation consisting of one object file for each object.

Figure 3.5

Infant Ordinal Choice



Feigenson carried out a number of controls that bolster confidence that successes reflected object file representations and failures reflected limits on the object file system. For example, it takes more time to put 3 crackers into a bucket than to put 2 crackers into a bucket, and the baby's name was used to draw attention to each cracker as it was put it. Perhaps infants were choosing the bucket that to which more attention was drawn or with which more time had been spent. In one control, on some trials the baby's name was called out and a hand waved over the bucket, but no cracker put in. Babies' choices were still determined by the actual number of crackers in the bucket, not the total time allocated to each bucket, the total attention drawn to each buck, or the naming events associated with each bucket. Conversely, perhaps the failures in the in the 3 vs. 6 condition was merely motivational. Perhaps once the number of crackers is so great, infants do not really care which set they obtain. To control for this possibility, an identical 3 vs. 6 condition was run except that the crackers were placed, one at a time, onto trays in full view. The infants overwhelmingly chose the tray with 6 crackers. I refer you to the original paper for these and other details that support the interpretation I offer here.

Consider 3 vs. 2 and 4 vs. 1. In both cases the total number of graham crackers was 5, and this number, by hypothesis, exceeds the upper limit of 3 infants can represent. Apparently, the limit is on attentional indices. Infants can create two short-term memory models of attended objects, up to the limits on parallel individuation, and compare them in memory. Be this as it may, the important lesson at present is this: the limits on set sizes of object tokens that may be simultaneously attended and tracked by adults are in the same range as the limits on the numbers of objects infants can simultaneously represent. This fact supports the identification of the representations underlying object-based attention in adults with the object representations of infancy.

3.7 Infant object representations: Occlusion vs. existence cessation

Another parallel between the two systems is that indexed objects, just like the objects represented by infants, survive occlusion. Scholl and Pylyshyn (1999) showed that object tracking in the MOT paradigm was not disrupted by the objects going behind real or virtual occluders. Almost all of the infant studies cited above involve occlusion—objects are hidden behind screens, in boxes, or in buckets.

In the MOT study of Scholl and Pylyshyn (1999) it mattered that the objects disappeared behind an occluder by regular deletion along its contour, reemerging from the other side by regular accretion along its opposite contour. If the objects disappeared at the same rate by shrinking to nothing, reappearing farther along the trajectory at the same rate by expanding from a point, tracking was totally disrupted. Similarly, if the objects just disappeared, all at once, reappearing all at once when they would have appeared on the other side of the occluder, tracking was totally disrupted. Thus, the system distinguished the object's going behind an occluder from its going out of existence, to be later replaced by another object coming into existence.

Berthenthal, Faith-Slaker and Kenny (2002) recently demonstrated that infants' object tracking is sensitive to exactly the same variations in how objects disappear at boundaries. They showed 5-, 7- and 9-month old infants a computer animation of a realistic 3D appearing rendering of a bright ball rolling along a track, disappearing behind a barrier and reappearing behind the other side. The dependent measures were the time spent fixating the event of disappearance and the degree of anticipatory looking for the ball to emerge out the other side. At 7 months (weakly) and 9 months (robustly) infants predictively tracked the balls that disappeared by continuous deletion along the edge more than those that disappeared instantaneously or by shrinking to a point. Conversely, at both 7 and 9 months of age, infants remained fixated for a longer time at the edge at which the ball was disappearing in the latter two cases than when the ball underwent continuous deletion at the edge.[6]

This pattern of results give us a third empirical argument in favor of identifying object representations of young infants with the representations that articulate adult mid-level object based attentional mechanisms. The infant's object tracking system and the object tracking system tapped in adult MOT studies use the same characteristics of events to distinguish two types of disappearance of currently attended objects: 1) disappearance that specifies continued existence of the objects behind the barriers and 2) disappearance that specifies existence cessation.

3.8 Conclusions from the identification of the two literatures

Researchers in both tradition, those studying infant object representations and those studying mid-level object based attentional mechanisms, have been studying the same

[6] There was one important difference in the pattern of findings in the infant study from those in the MOT studies: in the MOT studies the same results were obtained in the case of virtual occluders. This was not found for infants; infants did not predict the reemergence of the ball but rather fixated on the event of disappearance. This difference may reflect the fact that infants cannot yet automatically infer an occluder from patterns of disappearance, or, equally likely, they were trying to resolve conflicting data: the pattern of occlusion specified an occluder but there was no occluder visable.

natural kind. This discovery has important implications the characterization of core knowledge. First, the computations that establish object-indexes and object-files, that individuate and trace objects through time, operate throughout the life span, exemplifying one of the hypothesized properties of core knowledge. Second, object-file representations are the output of domain-specific, encapsulated, perceptual input analyzers, thus exemplifying another hypothesized characteristic of core knowledge. Adults may know that ducks do not change into rabbits, but the mid-level system that computes numerical identity in apparent motion studies does not use that knowledge. (This is what is meant by "encapsulation;" see Pylyshyn, 2001, for an extended discussion of the senses in which the processes of object-individuation and tracking in this mid-level system are encapsulated from conceptual knowledge of kinds). To a first approximation, the processes that compute figure-ground, assign surfaces to distinct objects, and assign indices to attended objects, work the same no matter whether the object picked out is a familiar kind or not (see Nakayama et al., date, for a review, but see also the work of Peterson, 2001, and Carey and Xu, 2001, for a caveat.) To repeat, this is why this system is called "mid-level;" the representations it computes are inbetween low level sensory representations and high level kind representations. Third, the capacity to establish object-file representations is innate, or at least emerges by 2-months of age (see Chapter 2), exemplifying yet another hypothesized property of core knowledge. I now turn to one final aspect of core knowledge systems—deep evolutionary history. Often, but not always, core knowledge is shared with other animals. This fact is important, for evidence that a knowledge system is shared among a wide range of species with a common ancestor but with very different ecological niches and different learning histories supports the hypothesis one should look to evolutionary pressures in our search for an explanation of its origin.

3.9 <u>The evolutionary history of object-file representations</u>

Chapter 2 described the work of Regolin and her colleagues on newborn chick's representations of spatiotemporal continuity of objects, representations that support search for occluded objects. This work provided an existence proof that representations of objects as spatiotemporally continuous may be innate, that is, may arise in the absence of any previous visual experience with occlusion. Do such results suggest that the mid-level object-file and object-indexing systems shared by human adults and young human infants have a long evolutionary history, perhaps arising early in vertebrate evolution? No, they don't. That chicks can form representations of objects that respect their spatiotemporal continuity does not warrant the conclusion that they have the whole system of core knowledge of objects. To explore this issue, we would need to characterize the conceptual role of chicks' object representations, whether these representations are subject to the solidity constraint and enter into computations of contact causality. We would also have to exp, chicks' object indexing capacities and limits on parallel individuation. My personal bet is that chickens do not share core knowledge of objects with humans, but it is an open empirical question.

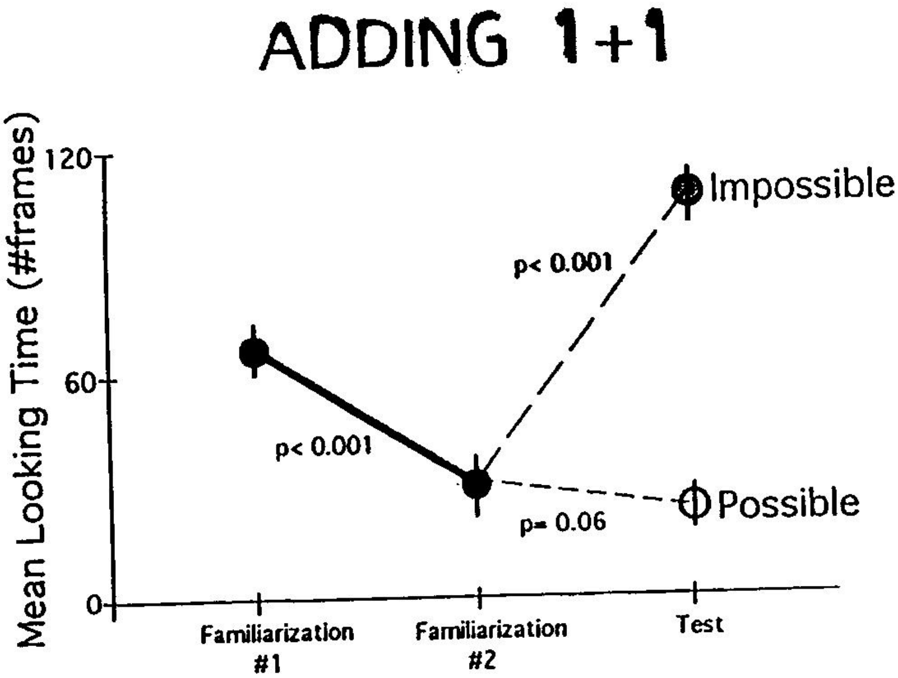
We do know that the evolutionary history of human core knowledge of objects extends at least into our primate past. Marc Hauser and his colleagues have used all of the methods reviewed in Chapters 2 and 3 (violation of expectancy looking time methodology, manual search for hidden objects, choice between two sets of hidden objects) with non-human primates (Hauser and Carey,1998; 2003; Hauser, Carey and

Hauser, 2000, Hauser, MacNeilage and Ware,1996). The results converge with the data from young infants in great detail. In addition to the theoretical importance of demonstrating the same system of core knowledge in non-human primates and preverbal human infants, Hauser's work has great methodological import as well. He was the first to show that the violation of expectancy looking time methods yield interpretable data with non-human primates, both free ranging rhesus macaques and laboratory housed new world monkeys, Cottontop tamarins.

Results from the violation of expectancy looking time methods show that both species of monkeys can use spatiotemporal evidence for object individuation and represent objects as continuing to exist when occluded. Consider an experiment with rhesus macaques on the island of Cayo Santiago in Puerto Rico (from Hauser, MacNeilage and Ware, 1996). This island supports a population of about 900 monkeys, in four social groups, who are subsist on natural vegetation and on monkey chow that is provided daily in feeding pens. Looking time studies can be carried out by identifying a monkey sitting calmly, alone (usually soon after feeding), approaching to within 10 feet or so, placing an experimental stage on the ground and carrying out the events. An assistant stands behind the experimenter with a video camera focused on monkey's face, and looking times are analyzed off-line by a coder who is blind to the experimental condition.

The first studies carried out using this methodology were two versions of a 1 + 1 = 2 or 1 study, and one version of a 2 – 1 = 2 or 1 study. I will describe one in some detail. Stimuli were eggplants, an unfamiliar food that kept the monkeys' interest. Because the monkeys are free to leave at any time (and often do), the experimental sessions were kept short. Each monkey was shown two familiarization trials, in which nothing impossible happened, and just one test trial with either an expected or unexpected outcome. Thus, unlike the infant studies, outcome (expected or unexpected) is a between subjects variable, and large numbers of subjects are run. Data from this study are depicted in Figure 3.6. There were a total of 72 monkeys, 24 in each of three different conditions. Monkeys looked less during the second familiarization trial than the first, indicating that they quickly became habituated to these events, and they showed no baseline preference for one vs. two eggplants. For the test trials, a screen was placed across the front of the stage and the monkeys were randomly placed into one of three conditions: 1 eggplant put in, 1 eggplant revealed when the screen was removed (possible outcome); 1 eggplant put in, followed by a 2nd eggplant put in, 2 eggplants revealed (possible outcome); 1 eggplant put in, followed by a 2nd eggplant put in, 1 eggplant revealed when the screen was removed (impossible outcome). As can be seen from Figure 3.6, the monkeys generalized their habituation to the two possible outcomes, and recovered interest when shown the impossible outcome. Similar results have been obtained with laboratory housed Cotton-top tamarins (Uller, Hauser, and Carey, 2001).

Figure 3.6.



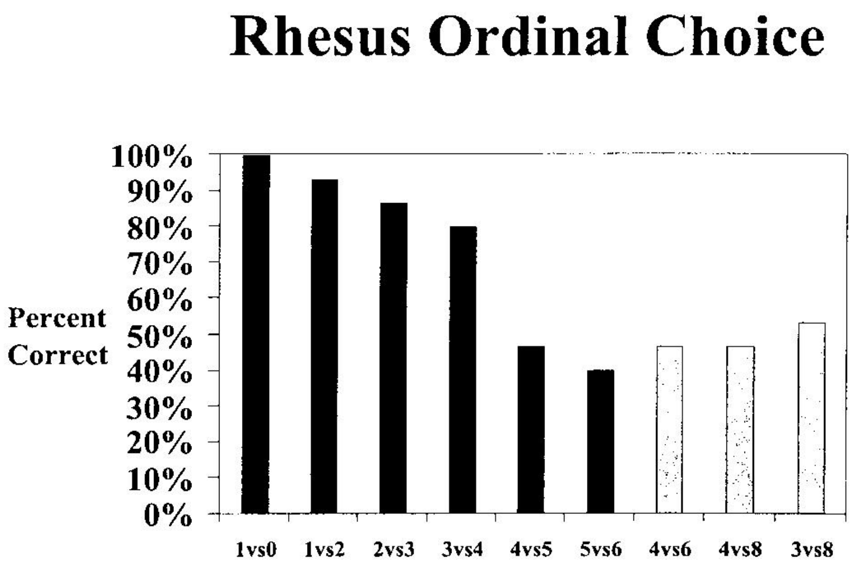
Hauser and Carey (2003) generalized the findings on Cayo Santiago to the same range of conditions under which success is obtained with 4 to 10-month-old human infants: 1 + 1 = 2 or 3 (showing that it is exactly 2 objects the monkeys expect; 2 + 1 = 2 or 3, 1 + 1 = 2 or big one (showing that monkeys are not solely encoding total expected eggplant volume.) Furthermore, performance breaks down at 3; monkeys fail at 2 + 2 = 4 or 3, consistent with there being an upper limit of 3 or 4 on the number of objects a monkey can track at once.[7]

Another paradigm that yields convergent results across the two subject populations is the set choice study described above (see Figure 3.5 and surrounding discussion). Actually, this paradigm was originally carried out on Cayo Santiago, with apple slices rather than graham crackers as the food item placed into the buckets (Hauser et al., ref). Monkeys watched as two experimenters placed one set of apple slices into one bucket, one at a time (e.g., 1 + 1 + 1), after which the other experimenter placed the another set into the other bucket (e.g., 1 + 1 or 1 + 1 + 1 + 1). Figure 3.7 shows the results; monkeys succeeded when the choices were 1 vs. 2, 2 vs. 3, and 3 vs. 4. Just like the babies, monkeys' performance fell apart when one of the sets exceeded a certain limit—in this case 4 rather than the 3 of the infants. Particularly important are the

[7] Hauser and Carey (2003) detail yet another way in which the monkey data converge with the infant data—in both cases, success is dependent upon the number of updates in short-term memory that are required to build a representation of the set behind the screen. Monkeys and babies succeed in 2 + 1 = 2 or 3 condition but fail at 1 + 1 + 1 = 2 or 3 conditions. Following Uller et al. (1999), we interpret these results as reflecting the processes that operate on short term memory representations of sets of object files.

failures at 4 vs. 8 and 3 vs. 8; these choices involve highly discriminable numbers, with ratios much greater ratios than those between small sets at which monkeys succeed (2 vs. 3 and 3 vs. 4). Again, the pattern of performance is extremely similar to that of the babies, and reveals the set size signature of object file representations (compare Figures 3.5 and 3.7).

Figure 3.7



Thus, primates reveal at least part of the core knowledge of objects that articulates infant object representations and object based attention in adults—objects may be individuated on the basis of spatio-temporal information, and models of small sets of objects may be stored in short term memory, supporting quantitative comparisons and action. There has been much less exploration of other aspects of the conceptual role of object-file representations—for example, solidity, contact causality, the distinction between coherent and non-coherent objects (see below). There has been one set of studies on tamarins' and rhesus' appreciation of solidity. To a first approximation, adults of both species resemble human infants in the solidity studies.

In sum, all the work to date suggests that the core knowledge of objects exhibited by young infants has a long evolutionary history. Cottontop tamarins, who last shared a common ancestor with human beings well over 100 million years ago, exhibit it, as do our more closely related cousins, rhesus macaques.

<u>3.10. What kind of *knowledge* is core knowledge</u>?

Object representations display all of the characteristic properties of core knowledge: the capacity to form them is innate, has a long evolutionary history, is supported by innate, encapsulated, perceptual input mechanisms, and persists throughout the life span. These are also properties of systems of perceptual representations. But an essential part of the core knowledge hypothesis is that the outputs of the perceptual input analyzers are *conceptual.* Core knowledge is conceptual knowledge. Chapter 2 detailed two respects in which representations of objects and contact causality have conceptual content: they cannot be defined in terms of spatiotemporal or sensory primitives, and they play a role in inferential processes that interrelate them with other representations that are conceptual in that same sense.

Some writers (including me, in my earliest writings on core knowledge) seem to be attributing infants conceptual representations in a much stronger sense (Baillargeon, 1993; Spelke, 1988). They speak of infants' "beliefs" that objects persist when occluded, infants' "knowledge" of that two objects cannot occupy the same place at the same time, infants' "reasoning" and "inferences" about the interactions of occluded objects, and their "surprise" or "puzzlement" at impossible events. There is nothing inherently wrong with such language, as long as one is clear what kinds of representations constitute the beliefs and knowledge in question, what kinds of computations constitute the reasoning and inference, and what kinds of states constitute the puzzlement and surprise. However, most researchers prefer not to use such highly cognitive language in describing the representations and computations of young infants, because it implies that the format of representation of the beliefs is propositional, and in that sense explicit, and that the computations carried out over such beliefs are logical inferences defined over propositions. While this may be true, most researchers who endorse the core knowledge hypothesis do not see it that way (see Carey and Spelke, 1994, 1996; Gelman, date; Leslie, 1994; Scholl and Leslie, 1999). Rather, they see the representations that articulate core knowledge as created by modular systems whose computations are constrained by the principles revealed by experiments such as those reviewed in these pages. For example, the computations that create representations of objects disappearing behind occluders by accretion along a boundary embody a commitment to spatiotemporal continuity. This is analogous to the sense in which the computations that create representations of depth from binocular disparity embody a commitment to different images in the two eyes arising from a single source. It would be decidedly odd to say that the 5-month-old infant "believes" that the images to the two eyes each derive from a single source in the world and thus provide information about depth, even though that infant undoubtedly uses that information to create representations of objects in depth (e.g., Held, Birch, and Gwiazda, 1980). I return to this issue below, when I consider the question of whether the representations in core knowledge are accessible or not, but to forshadow my argument, I do not believe that most of principles that constrain representations of physical objects are. Rather, I believe principles such as "one object cannot pass through the space occupied by another" to reflect constraints on the computations that establish representations of objects moving in the world. These are encapsulated compuations.

That the representations that articulate core knowledge are similar to perceptual representations in all these ways is important to the story I am telling in this book. Some concepts, those in core knowledge, have the content they do for the same reasons that some perceptual representations have the content they do—they are causally connected to the entities in the real world due to innate perceptual input analyzers that work as they do because natural selection has ensured it. But it is equally important to the story I am telling here that at least some of the representations that articulate core knowledge, the representations that are the output of modular input analyzers, the object-files themselves, have conceptual content.

Object-files are mid-level representations, in the middle between sensory representations and fully explicit conceptual representations of object kinds. Object-files are symbols for individuals in the world. If the object-file representations that articulate mid-level object-based attention and young infants' representations of objects in the world are created by dedicated perceptual input analyzers, perhaps we should think of them as *perceptual* symbols rather than *conceptual* symbols. This issue was the focus of Chapter 2, and it is worth revisiting, for it is central to our understanding of the nature of core knowledge.

That perceptual processes (figure/ground segregation, computations that establish the representations of surfaces, object-tracking on the basis of spatiotemporal information) establish object-files does not make them perceptual symbols, just as the fact that there are perceptual processes that identify instances of Michotte launching makes the representations of contact causality perceptual representations. Perceptual processes may deliver symbols that are conceptual, as seen by their content and conceptual roles. Chapter 2 argued that representations of *object* and *cause* are non-perceptual because their content cannot be stated in the vocabulary of sense data. We can do more, however, in addressing the question of whether object representations have conceptual content.

How do we decide what the content is of a mental symbol? It is difficult enough if the creature we are studying can express that symbol explicitly in language, but what of nonlinguistic creatures like chicks and monkeys and preverbal humans? I see no other route than studying what entities in the world cause the tokening of the mental symbol in question, that is, by studying the extension of the symbol. And also, we must study that symbol's inferential role. These two aspects of the functioning of symbols *determine* content, and thus what ever evidence we can glean about a symbol's extension and inferential role provides data relevant to discovering their content. What do infant object representations and adult object-files represent? I have argued that *object* is conceptual because it cannot be defined in terms of sensory or spatiotemporal primitives. But are we sure that the representations under study here actually have the content *object*? What entities in the world cause object-files to be established? How do these representations come to have real world objects in their extension? What computations do object-file symbols participate in? Do these computations provide evidence that object representations are interdefined with other representations with conceptual content?

<u>3.11. The extension of "object"-files and infant "object" representations..</u>

I shall argue that the content object-files is *physical objects*, by which I mean what is sometimes called "Spelke-objects," namely, bounded, coherent, 3-D, separable and moveable wholes. This claim, that object-files represent real 3D objects, hardly may seem surprising, but in fact, there are reasons to doubt it. In virtually all of the adult studies on mid-level vision, as well as in some of the infant studies (e.g., S. Johnson and Aslin, ref, M. Johnson and Gilmore, 1999, and Berthenthal et al., 2002), the stimuli are actually 2D entities on computer screens. Does the fact that 2D bounded entities activate object files mean that their content is more perceptual—perhaps "closed shape." No, it does not. We can present many of the cues for depth in 2D arrays, surfaces arrayed in 3D are routinely perceived in such displays. Furthermore, that the system can be fooled into accepting 2D entities as objects, does not mean that it is not representing the stimuli as Spelke-objects, just as the fact that the system can be fooled into seeing interactions among computer generated figures as launching events does not mean it is not representing such events as causal. But what reasons do we have for believing that the system is being fooled by these computer displays, and is representing them as real objects in spite of the fact that they are not? If 2D closed shapes are not in the extension of object-files, but object-files are activated by these computer displays, then the mid-level object tracking system is misrepresenting these stimuli.

Fodor (1998) has provided a way of thinking about misrepresentation. Misrepresentation is a general problem for any theory, such as his asymmetric dependence theory of referential content, that holds that content is determined by a causal link between the entities in the world and a symbol token in the mind. Fodor's example is a case where a horse seen in the distance on a misty day is misidentified as a cow. The real life horse caused the tokening of the *cow* symbol, and if content is determined by such causal links, doesn't this mean that the content of this symbol is horse as well as cow? In the present example, disks on computers causes an object-file to be opened. Doesn't this mean that computer disks are in the extension of object-files? No, says Fodor. When one fleshes out the causal story, one sees that the *reason* that the horse can cause the cow symbol to be activated depends upon the causal links between cows and that symbol, but that the reason cows cause the cow symbol to be activated does not depend in any way on whether horses do so.[8] This asymmetry allows us to see that the symbol really represents cows, not horses.

We could say the same thing about the relations between object-files and objects in the world. Representations that are part of core knowledge have the content they do because of causal links between entities in the world (objects in this case) and mental symbols (object-files). On the theory being proposed here, that 2D individuals cause object-files to be activated is dependent upon the causal relations that ensure that object-file refer to 3-D objects. In the case of core knowledge, natural selection has built perceptual analyzers that ensure the right causal paths between entities in the world and the symbols that represent them. It is clear how Fodor's theory allows that 2D entities might be misrepresented as objects, but we still need to provide arguments that this is the right analysis.

One reason to believe that infants misrepresent 2D pictures as real objects is that under at least some circumstances, they attempt to pick them up. One young toddler was observed trying to put his foot into a picture of a shoe.. Systematic studies have shown that attempts to handle and pick up pictured objects are readily elicited in children under one year old and disappear completely only around 18 months of age (see DeLoache, et al., 1998.)

Another line of evidence that 2D entities are actually being represented as objects is that the properties that govern their representations clearly reflect the properties of real objects. Section 3.7 presented one example. The processes that establish and maintain object-file representations are sensitive to the spatiotemporal information that specifies

[8] Of course, neither Fodor nor anybody else can actually flesh out the causal story. That is, we do not know how cows in the real world cause mental symbols for cows to be tokened, and thus we don't know how the concept cow has the content it does. But if we accept that there is some relevant causal story, then Fodor's assumptions seem warranted.

either occlusion or existence cessation. Occlusion and existence cessation are properties of real physical objects, not computer disks.

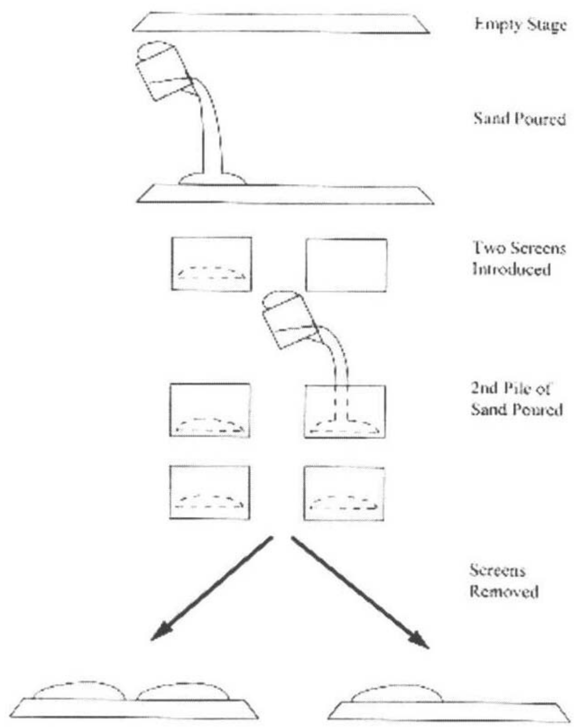
The implosion/disappearance provide data concerning the extension a mental representation by probing what does *not* cause its tokening—in this case entities that shrink to nothing or suddenly disappear fail to elicit object indexing and tracking. Two series of studies with 8-month-olds infants confirm that mere perceptual boundedness is *not* sufficient to cause object-files to be set up.   The individuals being tracked in the infant studies are physical objects, and not just any perceptual objects specified by figure/ground processes, such as disks on computer screens or piles of sand or blocks.

.A hallmark of physical objects is that they maintain their boundaries through time. Neither a pile of sand nor a pile of blocks is a Spelke-object, in spite of the fact that when stationary each may be perceptually indistinguishable from one. One may make a pile-shaped cone and coat it with sand, or one may glue together a pile of objects, yielding a single pile-shaped entity. It is only upon viewing such entities in motion (do they fall apart, or do they maintain their boundaries?) that unequivocal evidence for their ontological status is obtained.   Consistent with the claim that object-files represent *objects*, two series of studies establish that infants track Spelke-objects that are perceptually identical to piles of sand (Huntley-Fenner, Carey, and Solimando, 2002) or piles of little blocks (Chiang and Wynn, 2000), under conditions where they will not track the perceptually identical non-objects.

Take Huntley-Fenner et al. (2002) for example. Huntley-Fenner carried out a series of 1 + 1 = 2 or 1 studies involving sand poured behind a screen or sand-objects being lowered behind a screen.   Stimulus type was a between-participant variable, and infants were familiarized with the stimuli before the study, by handling the sand or the sand object. One study involved a single screen; another involved two screens.. Eight-month-old infants succeeded in the sand-object conditions, but failed in the sand conditions.

The failure in the two screen study is especially striking, for it shows that infants do not have “sand permanence.” In this study, diagrammed in Figure 3.8, the infant watched as a pile of sand was poured onto the stage floor, and then covered by a screen. A second, spatially separate, screen was introduced and a second pile of sand poured behind it. The screens were then removed, revealing either two piles of sand (one behind each screen) or only one (the original pile initially seen on the stage floor). Eight-month-olds did not differentiate the two outcomes, although they succeeded if the stimuli were sand-pile shaped Spelke-objects lowered as a whole onto the stage floor. To succeed at this task, the infant need only represent “sand behind this screen, sand behind that screen.” Why did they fail at sand permanence? As belabored in Chapter 2, object permanence requires an individual whose identity is being tracked; it is the *same individual* we represent behind the screen. Apparently, 8-month-old infants cannot establish representations of individual portions of sand and trace them through time. Chiang and Wynn (2002) found exactly the same results with piles of blocks. If the pile moved as a single Spelke-object, 8-month-old infants could track it and represent it as continuing to exist behind a barrier. If they were shown this “object” be separated into 5 blocks and then reassembled, they subsequently failed to track it.

Figure 3.8



Our interpretation of these findings is that piles of sand and piles of blocks, being non-coherent, fail to activate the object-file system. These infant studies suggest that the object tracking system is just that: an *object* tracking system, where *object* means three-dimensional, bounded, coherent physical object. The object tracking system fails to track perceptually specified figures that have a history of non-cohesion.

Let us stop and take stock of where we are. Representations of real-world, 3D, spatiotemporally continuous objects are not perceptual because they cannot be specified in terms of perceptual primitives. But in order to be sure that infants have mental symbols with the content *object* we must study the actual real world entities that cause the tokening of the mental representations we believe have that content—the mental representations that underlie infant performance in the experiments reviewed in Chapters 2 and 3. The sand and block-pile studies reviewed above add data in support of the claim that the content of what we are calling *object* representations are indeed real world objects. The other relevant data come from studies of the inferential role of infant object representations. I now turn to inferential role, amplifying the arguments offered in Chapter 2.

3.12 The inferential role of "object-files" and infant "object" representations

There has been no work on the inferential role of adults' object-file representations, but if we accept the identification of adult object-files with infant object representations, then the infant work bears on both. Chapter 2 gave several examples of inferential interrelations of young infants' object representations with other representations that go beyond perceptual content. That is, object representations articulate physical knowledge. As Chapter 2 showed, by 2 months of age infant object-file representations are quite adult-like. Infants of this age are sensitive to almost all the same information adults are in building representations of objects that amodally complete surfaces behind barriers, although young infants need more redundant cues than do older children or adults (Scott Johnson, date). Astoundingly, 2-month-olds are also able to represent physical relations such as inside and behind, and their representations are constrained by knowledge of solidity, a property of Spelke-objects but not of 2D visual objects (Hespos and Baillargeon, date, Spelke et al, 1992.) Besides expecting objects to be solid, and thus not to pass through other ones, slightly older infants (6-month-olds) also expect objects to be subject to the laws of contact causality (e.g., Leslie, 1988; Leslie and Keeble, date).

Thus, the conceptual role of the infant's object representations is that of 3D Spelke-objects; objects are represented as solid entities in spatial relations with each other, that cannot pass through other objects, and which move only upon contact. If we accept the identification of the infant's object concept with object-files, then we must accept that object-file representations also have the same conceptual role. Object-file representations have conceptual content not only because they cannot be reduced to sensory primitives but also because they are inferentially interrelated with other representations that cannot be reduced to sensory primitives.

3.13. Modularity: encapsulated or informationally promiscuous?

As discussed in Chapter 1, Fodor (1983) suggested that rather than trying to distinguish perceptual from conceptual content, the conceptual/perceptual contrast is more perspicuously drawn on the basic of processing characteristics. He suggested that perceptual processes are modular, and he characterized modular processes as fast, automatic, primarily data driven by sharply limited input, inaccessible, and encapsulated. By encapsulated, he meant that other knowledge does not affect processes internal to the module. He contrasted modular perceptual processes with central cognitive processes, which are slow, effortful, optional, accessible, and informationally promiscuous. By that, he meant that there are no restrictions on what data bear on which inferences; it is a matter of theory building to discover the inferential relations among real world phenomena. Notice that if one draws the distinction between perceptual/conceptual this way, many representations that do not have sensory content turn out to be perceptual, such as syntactic representations.

Modular perceptual input devices create representations of Michotte causality and individuated 3D objects. Although the output of these processes have conceptual content in the sense of not being statable in terms of perceptual primitives, like syntactic representations, these representations would seem to be perceptual on Fodor's definition. The core knowledge thesis concurs that core knowledge representations are perceptual in this sense. Indeed, it is important to the thesis, for the existence of evolutionarily created innate perceptual input analyzers at least partially solves the problem of how the representations in core knowledge have the content they do. This aspect of core knowledge representations explains how they are causally connected to the entities in the world they represent.

However, I do not accept that core knowledge representations are perceptual even by Fodor's characterization. At issue is whether the representations that are the *output* of the innate perceptual input analyzer are part of a central system that is cognitive by Fodor's characterization. Is an object-file itself or launching causality itself accessible and do such representations participate in slow, optional, inferentially promiscuous processes? Are they inferentially related to the outputs of other systems of core knowledge? I believe all these questions should be answered in the affirmative, and I shall try to convince you of that. Most important, though, is that this is what the question of perceptual/conceptual comes to on Fodor's analysis.

Take first the question of accessibility. Of course, it is virtually impossible to know whether a representation in a prelinguistic creature's mind is widely accessible. But object-files are accessible. We have phenomenal access to them and we can carry out a wide variety of optional computations, under executive control, over them. Accepting the identification of object-files with infant object representations implies that object-files are accessible for infants as well. We can do better. Object files support voluntary action—infants reach for objects, even hidden ones. The box search and bucket choice studies reviewed above show that infant object representations, like object-files are individual symbols that can be placed in short term memory, and such short term memory representations are accessible for adults.

What clinches the matter for me is evidence that infant object representations interact inferentially with representations that are the output of distinct input modules. Individuating distinct domains of core knowledge is far from a trivial matter. It is not clear, for example, whether computations of Michotte causality are part of the object module or a separate system whose outputs are interrelated with it. But on just about every analysis, spatial representations are a distinct input system from object file representations, as are number representations, quantity representations, and representations of intentional agency. Yet object representations are integrated with representations in all of these domains.

With respect to spatial representations, infants represent the spatial relations among object files. They represent objects behind barriers, inside boxes, and inside buckets. In one spectacular series of studies, Onishi and Baillargeon (date) habituated infants to two identical blocks being moved into the center of the stage from the sides, with one placed on top of the other. Looking times are measured to the static array. After habituation, the blocks are moved to the center of the stage, but the previously botton block is placed on top. The resulting array is identical in appearance to that the infant was habituated to, but if they distinguished the two object tokens and represented which one was on top, their attention might be drawn to the change, and indeed, it was.

With respect to number representations, infants can compute 1-1 correspondence over object-files to establish numerical equivalence (Feigenson and Carey, 2003). With respect to quantity representations, infants can sum over continuous variables bound in object files to choose between sets on the basis of total volume (Feigenson et al, 2002a). And with respect to agency, infants as young as 5 months of age represent objects as goals for of others' intentional action (see Chapter 4 below). Even more strikingly, 7- to 12-month-old infants infer a previously unseen agent to explain the motion of a known inanimate object (Saxe, Tenenbaum, and Carey, under review). For example, in one condition infants were habituated to a beanbag flying out of one of two boxes and landing on the stage between them. After habituation, the fronts of the boxes were lowered, revealing a hand either in the box from which the beanbag had emerged or in the other box; 7- and 10-month-old infants looked longer when the hand was revealed in the other object.

Thus, while the computations that yield representations of object files are modular and encapsulated, the object-files themselves are inputs to a variety of central computations. With respect to this aspect of inferential role, then, the object representations that are part of core knowledge are conceptual.

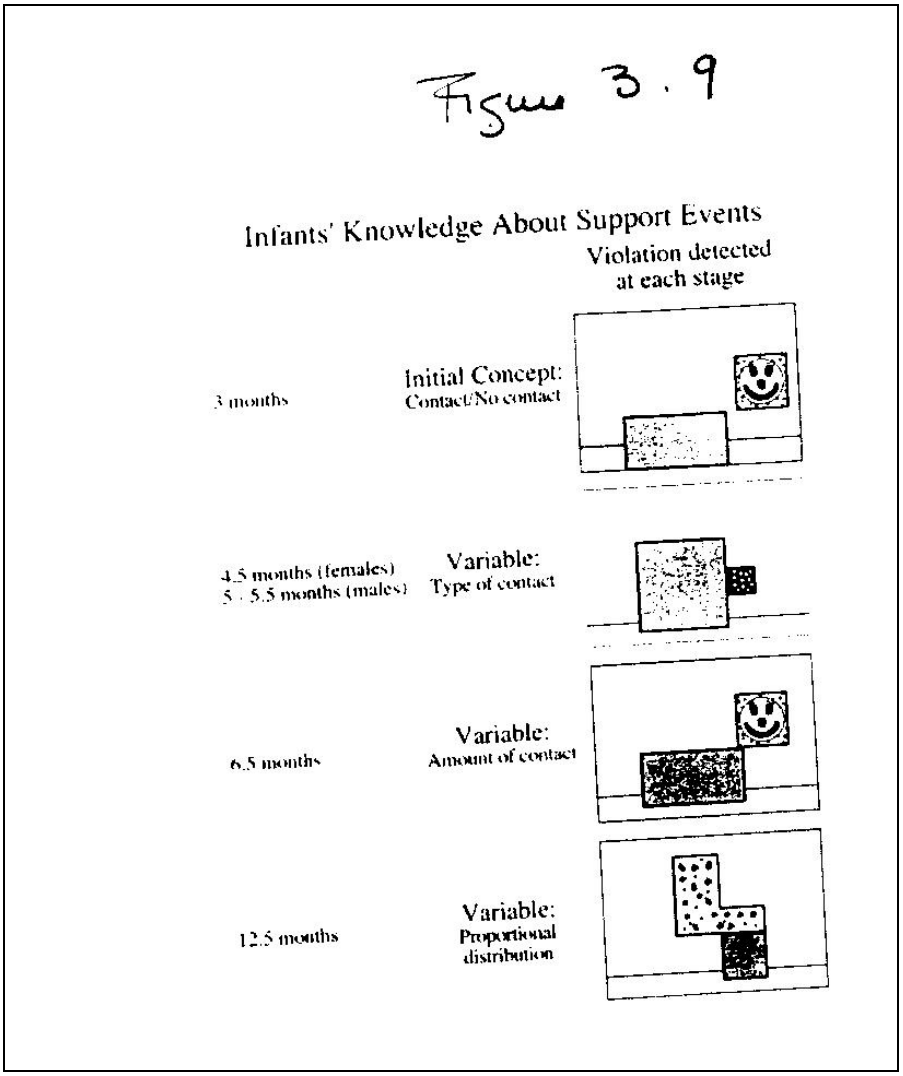
Notice that on this view, some of the representations that articulate core knowledge are not conceptual—those that are within-module and encapsulated. I have made an extended argument that *object* itself is a conceptual representation, but knowledge of spatiotemporal continuity and cohesion probably are not. The computations that create representations of objects make use of evidence for spatiotemporal continuity and boundedness, and embody a commitment to these properties of objects in further computations, but there is no reason to believe that the child explicitly knows principles such as “objects continue to exist behind barriers, or objects do not fall apart and reassemble.”

3.14 <u>Core knowledge and learning.</u>

Up to now we have considered the evidence that core knowledge representations are the output of innate perceptual input systems, that these continue to function throughout the life span, and that the outputs of these modular devices are conceptual. Another hypothesized feature of core knowledge systems is that they are learning devices. There is no doubt that infants learn many generalizations about objects during their early months. Thus, the processes that yield object representations yield representations about which the infant learns. What we do not yet know is to what extent the learning processes are within module and domain specific or general central processes. While this is an important question, either possibility is consistent with the core knowledge position.

Consider just one example. Infants do not innately know that unsupported objects fall (Baillargeon, 1998). In one series of studies, infants watched while a small block was slowly pushed across a large supporting block (Figure 3.9). They were habituated to the display on the top, in which the block began on one end of the support and ended at the other, still fully supported. Then infants were presented a series of test trials, probing their expectations concerning when the block should fall (A – F). On some trials the small block was pushed off the large one until it was completely unconnected, and thus totally unsupported by it, apparently suspended in mid air (A). On other trials, it remained in contact with the block, but in various configurations that to adults either would (D and F) or would not (B, C, and E) continue to support the block.

Infants’ expectations unfolded in a regular sequence over a long period of development. At 3-months of age, infants showed no differential interest in these these events. Even the unsupported object (A), hanging in mid-air, was not particularly attention grabbing. Just a few weeks later, though, the impossible event draws markedly longer looking than does the possible event they were familiarized to. The child has begun to learn something about support.



In a series of beautiful experiments, Baillargeon has shown that infants' learning about support unfolds in a regular way. First they make a categorical distinction between contact/non-contact, and do not pay differential attention to objects that do not fall so long as there is any contact with the support. That is, they look longer at the outcome in Figure outcome A in Figure 3.9 than at any other outcome, including B, but do not differentiate any of the others. They gradually refine the parameters relevant to support. Next the contact must be from below (now outcome B also draws attention, but none of the others do). Then more than half of the base of the object must be supported from below (C also draws attention, but none of the others do. Finally, they take into account the geometry of the object (Figure E is attention grabbing but not D or F).

Baillargeon and her colleagues have indirect evidence that the initial stages of this learning occur, in the ordinary course of events, from infants' own attempts to place objects on surfaces. Infants who sit unsupported progress through the early steps of this sequence earlier than those that do not yet sit alone, consistent with the hypothesis that infants learn about support by placing objects on surfaces and observing the outcomes. Those who sit alone have their hands free to manipulate objects. Baillargeon has shown that learning about support can also be driven from observational evidence. In training experiments she shows infants contrasting cases of objects being placed on surfaces and falling or remaining supported and finds acceleration in the above sequence.

The objects involved in the support studies are unfamiliar to the babies; that is, they have not had experience with those very objects. This suggests that their previous experiences with objects in general, with Spelke-objects, is driving the developmental progressions Baillargeon observes in these studies. This point is important and merits further study, for it bears on the domain-specificity of these learning mechanisms.

3.15 Domain general or domain specific learning mechanisms?

Experiments such as Baillargeon's certainly show that infants learn about objects, but they leave open whether the processes that support this learning are at least partly domain specific. To my knowledge, there has of yet been no systematic study of this question. What kind of domain specificity might we expect to find, how could we look for it, and what would the core knowledge position predict?

It is easy to see how the learning from observation in the support studies could be well modeled by domain general associative mechanisms that extract statistical regularities from representations of events. The sense in which domain-specific learning mechanisms *may* be involved is limited, but important. There may be domain-specific constraints on the features that enter into the statistical process.

An analogy from the animal learning literature clarifies the sense of domain-specific learning at issue here. There is absolutely no doubt that animals learn associations between stimuli. Rats can easily learn that the occurrence of a particular sound predicts the occurrence of a shock from their water feeder, and they can also learn that a distinctive taste in the water predicts nausea 2 hours later. Although this is true, there are also constraints on the associations animals learn easily. That is, the above pairing is easy to learn; the reverse (that a sound predicts nausea, that a taste predicts shock) is much harder to learn. There are domain specific constraints on the associative pairings that can be learned. The appeal to domain specificity in this example is much weaker than in the case of the domain specific mechanism that enables Indigo Buntings to extract north from rotation of the night sky. This latter mechanism specifies a particular computation on the input that generates the required representation. In the rat case, the associative mechanisms are very general, applying to a huge variety of cases of learning that involve computing statistical covariation in the environment; the domain specificity comes in constraints on the salience weighting of particular features in particular contexts. Although weaker, nonetheless, this is a bona fide type of domain specific constraints on learning (see Gallistel et al., 1991, for an extended discussion of species specific and domain specific constraints on associative learning).

How would we find out whether the processes that extract the statistical generalizations concerning support are domain specific in the sense of being constrained to weight some features more heavily than others? It is possible to imagine a relevant program of research. For example, one could take a variety of contrasts among objects that are salient to infants—e.g., shape contrasts—and provide statistical evidence that these covary with whether objects remain supported or fall in the observational learning paradigm of Baillargeon and her colleagues. That is, one could try to teach the generalization—a cylinder covered with blue glitter, supported from below on ¾ of its surface, does not fall; but a red striped block, supported from below on only ¼ of its surface, does fall. What generalization does the child learn—the geometric one concerning amount of surface supported, that cylinders don't fall and blocks do, that blue glittery things don't fall but blocks do? If the child is biased to analyze the geometric relations between the object's base and the support, this would be evidence for domain specificity in this learning mechanism. The core knowledge position requires that at least some such constraints on statistical learning will be observed, for, by hypothesis, core knowledge systems are domain-specific learning devices.

3.16. Challenges to the core knowledge hypothesis.

It is an understatement to say that not all students of the infant mind agree with the characterization of the initial state I develop here. The hypothesis that young infants' representational capacities include several domains of core knowledge, as core knowledge is characterized in these pages, is highly controversial. As mentioned in the introduction, some of the controversy derives from differences in scientific taste, and people prefer leaner interpretations of the data I have been offering in favor of core knowledge of objects. But scientific taste cannot be the arbitrer of truth. Ultimately, the question is empirical. Do the data favor the core knowledge hypothesis over the Empiricist alternative? One goal of this book is to display the empirical case in support of core knowledge, which I believe to be overwhelming.

Indeed, challenges to the core knowledge position have often been empirical. Many of these focus on the looking time studies that are the main source of evidence concerning young infants' mental representations. Because the general points the skeptics raise are excellent ones, it is worth laying them out and thinking about them. The criticisms boil down to two observations that are undoubtedly right: First, infants' looking patterns are determined by many factors in addition to responses to violations of expectancy, and it is difficult to rule out alternative interpretations of the patterns of looking obtained in a given experiment. Second, the very late emergence of some explicit representations (representations that can guide action, for example) of knowledge putatively contained within core knowledge raises questions about the very existence of core knowledge. Let me take up these two objections, in turn, focusing on knowledge of solidity as my example.

3.17 Alternative explanations of the looking time patterns in the solidity experiments.

The first objection to inferences from looking time studies draws on the observation that infants' attention is drawn not only by violated expectancies, but also by simple perceptual novelty or familiarity, and even more basically, by intrinsic perceptual preferences. Thus, any looking time study must control for these latter bases of looking adequately before patterns of attention may be taken as reflecting sensitivity to a violated expectancy. The critics claim that the controls are not adequate, and that the observed patterns of looking reflect simple perceptual preferences or familiarity/novelty preferences.

Take the solidity experiments by Spelke et al., 1992, for example. In those experiments, 4-month-old infants were habituated to an object's being lowered behind a screen, after which the screen was removed revealing the object on the stage floor (Figure 3.10.). In the test trials, a solid shelf was introduced into the apparatus, the screen replaced, and the object lowered, as before, behind the screen. The screen was then removed, revealing either a possible outcome of the object resting on the shelf (A), or an the impossible outcome of the object resting on the stage floor as before, apparently having passed through the solid barrier (B). Infants looked longer at the impossible outcomes, consistent with the suggestion that knowledge that one object cannot pass through the space occupied by another constrains the models the infants build of events unfolding before them.

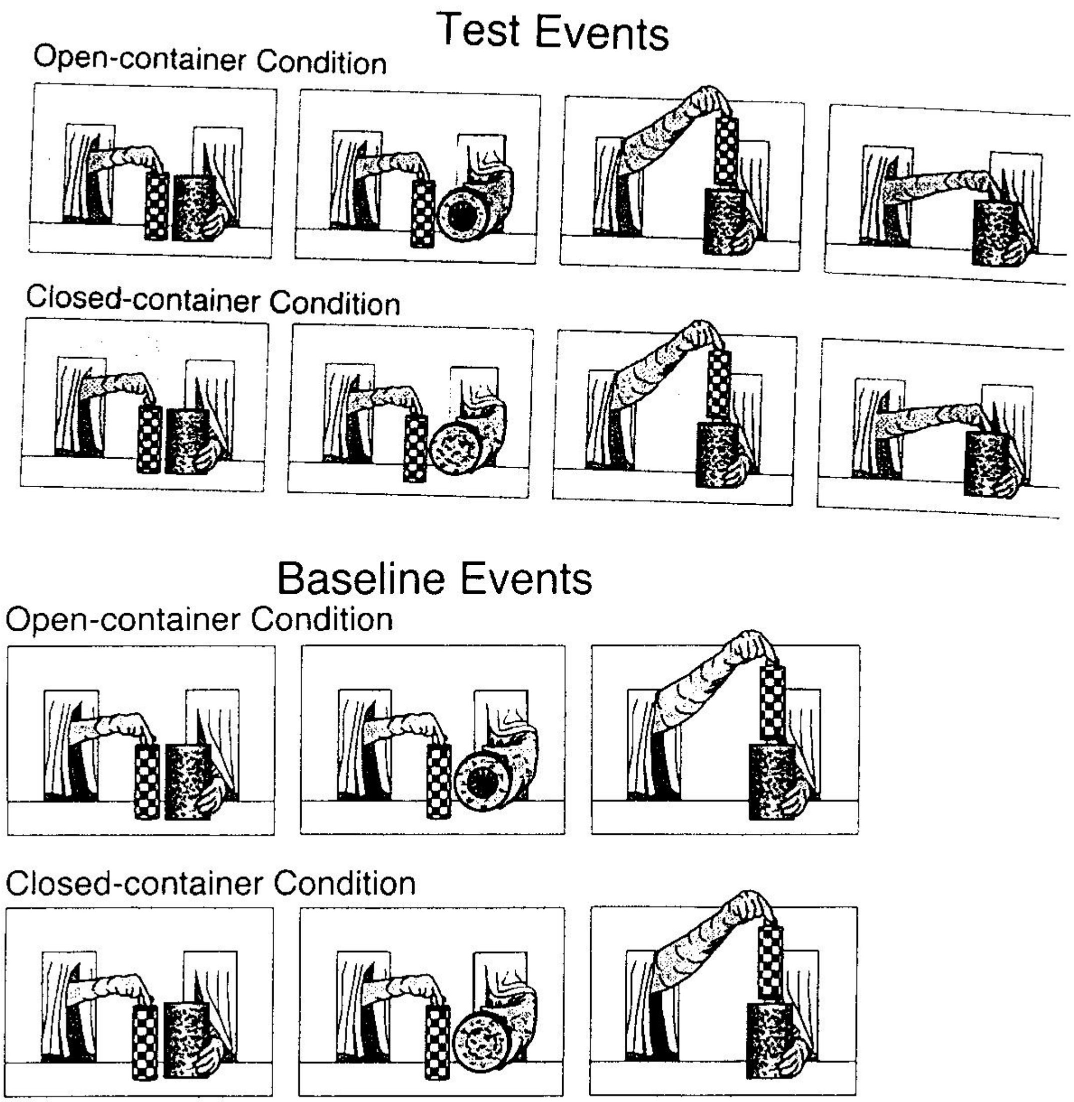
Critics of these studies have suggested that the outcome in which the object appears enclosed (between the shelf and the apparatus floor) may be more visually interesting to infants than the outcome in which the object is outside, perhaps because of greater contour density (e.g., Haith et al., date.) Of course this is possible, but all experimenters using these techniques are aware of such possible confounds and take care to eliminate them. Often, baseline conditions involve showing infants just the outcomes of a series of test trials, to establish whether there are intrinsic preferences. In a replication with 8-month-olds, Huntley-Fenner et al. (under review) found no baseline preference for either outcome, in the face of a large preference for the impossible when infants saw an object lowered onto a shelf behind a screen, the screen then being removed revealing either the object on top of the shelf (possible) or below the shelf (impossible).

A slightly different line of criticism relies on novelty preferences rather than intrinsic preferences. Remember, infants were first habituated to the object being lowered onto the stage floor, after which the barrier was introduced for the test trials. Spelke et al. reasoned that the impossible outcome is actually perceptually more similar to the familiarization outcome (because the object is in exactly the same position), so that a novelty preference would support the opposite results to those obtained. Critics have replied that it may as well be just the opposite—an object resting on the first surface reached from the direction from which the ball has come may be more perceptually similar to the familiarization outcome, and thus the preference for the impossible outcome may merely be a novelty preference. Again, although this is logically possible, it is possible to control for, and Spelke et al. (1992) did so. They included a condition in which the object was placed in position by hand either on the shelf or below the shelf. The familiarization/test sequences were exactly the same as in the solidity experiments (i.e., several familiarizations with the object having been placed on the stage floor, the barrier being introduced, and then alternating test trials in which the object is either placed on the shelf or on the stage floor). In this case, there was no preference for the one outcome over another, ruling out a novelty preference for the impossible outcome in the experiments in which the motion of the object would have violated the solidity constraint.

Although researchers using the violation of expectancy paradigm attempt to rule out perceptual preference or perceptual novelty artifacts, it is always possible that somebody will find one that accounts for one, or even many, results in the literature. Looking times are sensitive to such factors. But in evaluating the evidence for young infants' representations of object motion being constrained by the principle that one object cannot pass through the space occupied by another, one must consider the huge amount of convergent evidence from many different paradigms (consider the Baillargeon screen studies; the Baillargeon box on track experiments, the Baillargeon sand in cylinder experiments, the Hespos and Baillargeon rod/cylinder experiments, the Spelke et al. shelf experiments). Convergent evidence from many different studies, each with different controls for perceptual preferences or perceptual novelty, ultimately convinces me of the case.

Although I will not belabor the controls in each study I cite, I would like to reassure the skeptical reader by going through one particularly elegant case—the Hespos and Baillargeon (date) solidity study. See Figure 3.11 for a diagram of the test events in this study. Recall that 2-month-olds looked longer at an impossible event in which a rod apparently was lowered through a remembered lid on a container than when it was lowered into an empty container. This study scrupulously controlled for novelty and familiarity effects. To flesh out the full design, infants were habituated to the rod being picked up and raised over the cylinder, as in Figure 3.12. This familiarized the infants with the hand, the rod, the motion of the rod, and so forth, but provided no information about the upcoming test events. After familiarization, in a between participants design, the cylinder was turned revealing either a closed top (1/2 of the participants) or an open tube (1/2 of the participants), and then it was turned upright again. Then the rod was picked up, raised over the cylinder as before, and then inserted into the cylinder (impossible outcome for those participants shown the closed top; possible outcome for the other group; these are the test trials depicted in Figure 3.11). Those for whom the event violated the solidity constraint looked over twice as much as did those for whom it did not.

Figures 3.11, 3.12



Consider this design. Both groups saw exactly the same events during familiarization and during test; the only difference was what was revealed when the cylinder was turned. Since there was only one outcome during the test trials (the rod was lowered into the container), a simple perceptual preference for one outcome over another could not have accounted for the results. Since the habituation events in both conditions were identical (the rod was lifted and held above the container), greater perceptual novelty of one of the outcome events relative to the habituation events could not account for the results. A final possibility is that the closed top was more interesting than the open tube, and so infants looked longer during the test events in the closed top condition just because of that. Hespos and Baillargeon controlled for this as well, running two more groups in a series of test trials that were identical to the familiarization trials (the rod remained perched above the cylinder). Between the familiarization and test trials, ½ of the participants were exposed to the open container and ½ to the closed. There was no difference between the groups; having seen a closed tube did not by itself cause greater interest in the events. As far as I can see, this leaves no explanation for the pattern of looking other than the violation of the solidity constraint.

It is beyond the scope of this book to include all the controls for perceptual preferences and perceptual novelty/familiarization effects that have been included in the experiments I report. I advise the interested or skeptical reader to look at the original papers I cite. I acknowledge that it is very important to worry about such effects, for they surely influence the observed patterns of looking times in these studies. Nonetheless, I am convinced that such low level effects cannot account for all the data. Basically, the looking time studies always include relevant controls. This is not to say that the controls have always been adequate. There are no guarantees in science. No doubt, the interpretations of some particular studies will be modified in the future, as alternative interpretations of the observed patterns of looking times are corroborated. Nonetheless, the conclusions I draw from the experiments cited in these pages are supported by a great deal of convergent evidence from many different paradigms.

3.6.1 Looking time vs. reaching

As we saw in Chapter 2, infants' failure to reach for hidden objects has led many researchers to doubt that young infants actually do represent objects as existing behind barriers. Chapter 2 laid out a response to these doubts. Performance limitations, due to immature executive function (Diamond, 1991) or to parameters of the motor system that guides the reach (Thelen et al., 2000) partially account for failures to search in such tasks. Development within these systems can be independent of the capacity to form representations of hidden objects. The worry from these developmental lags is mitigated by the fact that the age differences in success in the looking time procedures and the search procedures, when the question at hand are the individuation and tracking of hidden objects, is often a matter of just a few months.[9]

However, sometimes the developmental lag is a matter of years, casting doubt on the infant studies or at the very least requiring new ways to think about the discrepancy between knowledge revealed by looking time methods and knowledge revealed by explicit measures of prediction, search, and pointing. The solidity studies provide a particularly striking example of much earlier manifestation of understanding in looking time procedures than in procedures that rely on search. Two-month-olds succeed in the looking time studies (Hespos and Baillargeon, Spelke et al, 1992), whereas it is not until

---

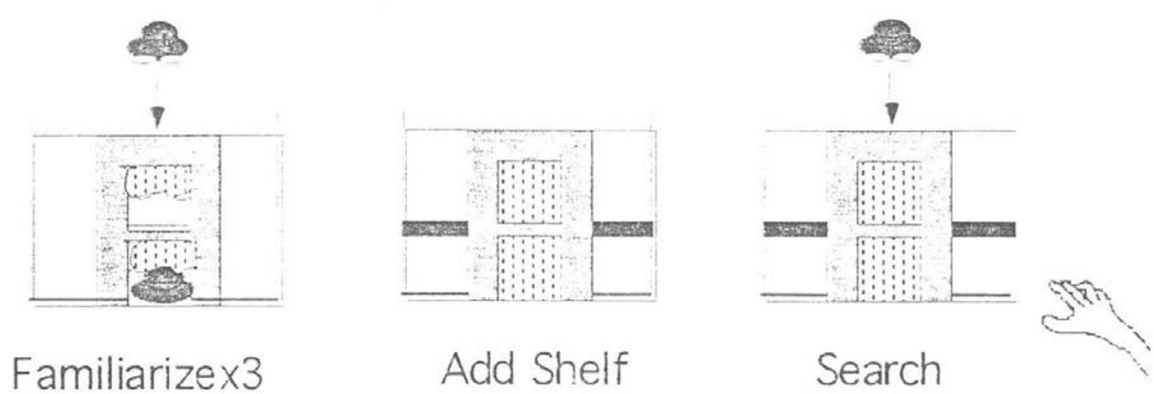
[9] Sometimes there is no difference at all. Compare Xu and Carey, 1996, a looking time study, and Van de Walle, et al., 2000, a manual search procedure. In both cases, 10-month-olds fail to use property/kind information to individuate objects and 12-month-olds succeed.

over *2 to 3 years* of age that children succeed in manual search versions of the same tasks (Hood, Carey and Prasada, 2001, and Rochat and Clifton, 1999).

Hood et al. carried out four studies closely modeled on the Spelke et al. shelf studies, and in all four cases, 24-month-old infants failed. In the studies of Hood and his colleagues, 30 month-olds succeed, whereas in Rochat's and Clifton's studies, robust success is not obtained until 36 months of age. Take one of Hood's studies as an example. Toddlers were familiarized with an object being dropped onto a stage floor. Then a barrier was introduced above the stage, and a screen with two doors, one at barrier level and one at floor level, was introduced in front of the apparatus (Figure 3.13). The object was dropped behind the screen, and the infant was allowed to search for the object. Two-year-olds searched at the floor level, where they had seen the object before, and not at the barrier level, where the object must be if its motion was arrested by the barrier.

Rochat's and Clifton's studies are all the more striking. Their task is more difficult, involving four doors, and four possible locations of the barrier (Figure 3.14). But unlike Hood et al, they gave the child considerable training. The child watched the ball roll down the incline with the doors open and the barrier placed at each of the four locations, seeing it stop in front of the barrier each time. Also, they were given multiple trials, with feedback. As in Hood et al.'s studies, toddlers do not begin to succeed until between 2 ½- and 3-years of age.

Figure 3.13



A first response to these studies is to doubt the infant work. How can 2-month-olds know that one object cannot pass through another when 2 year olds search for an object as if it had fallen right through a solid shelf? Some of the explanations offered in Chapter 2 for the A/not B error in search tasks probably play a role in this case as well. The toddler's search shows the influence on search of previously viewed locations of the object. Part of the answer lies in developing executive function, which must adjudicate between competing representations of locations. Perhaps there is further development of the system of planning that computes over representations in motor work space. But I do not think this is the whole story, because of the length of the gap. The difference between 2 months and 3 years suggests a principled difference in the kinds of knowledge being tapped in these two different paradigms. Indeed, I believe there are two kinds of knowledge involved here: within-module encapsulated representations and explicit representations that are output of the perceptual device that creates representations of object files. A related qualitative distinction that applies to this case is that between prospective vs. retrospective modeling processes (Berthenthal, date).

Solidity constraints on object motion follow from cohesion and spatiotemporal continuity, which, I argued above (Section 3.13), are within module constraints on representations, and not themselves explicitly represented. There is no reason to believe that the infant has access to a generalization such as "one object cannot pass through the space occupied by another." Indeed, the toddler data provide interesting evidence for the distinction between accessible representations within core knowledge and encapsulated ones. Explicit representations of objects existing behind the screen are the output of the perceptual analyzers that create object representations (note that the toddlers do search for the object, just as 8-month-olds do), but explicit representations of where the object should be relative to the barrier are not.

Retrospective and prospective modeling processes may draw on encapsulated and explicit representations, respectively. In violation of expectancy looking time studies, the perceptual system need only check whether a seen outcome is consistent with the representation of previous events. Apparently, when the outcome of the event is revealed, the infant can check it for consistency with a memory representation of the object's initial location and path of motion. If within-module constraints on motion are violated, attention is drawn, for the process that builds a consistent model of the whole event makes use of all the information available and embodies the constraints that operate in the domain. Of course, if the explicit representations are violated, attention is drawn as well. But what the child cannot do is use a representation of the path and of the now hidden barrier to form an explicit, prospective, representation of where the object will be relative to the barrier. On the modularity view, this is because knowledge of solidity is not explicit.

The striking failures of the toddler do not undermine the conclusions from the success of very young infants, but they do reinforce the need to be cautious about what kind of knowledge we attribute to them. It seems likely that knowledge of spatiotemporal continuity, of the boundedness and cohesion of objects, and of the solidity constraint, is implicit in the computations that create models of objects and their interactions. As Pylyshyn (2001) emphasizes, this knowledge is encapsulated within the input module that creates representations of objects, even if the symbols for the objects themselves, the object-files, are explicit and conceptual.

4 Conclusions

The infants' knowledge of objects displays all of the hypothesized properties of core knowledge. Representations of objects are created by modular, encapsulated, perceptual input analyzers, and these representations continue to articulate our representations of the world throughout life (modulo within-domain learning). Object-files are conceptual representations in that they cannot be defined in perceptual primitives, and as shown by their rich conceptual role and their accessibility. The system of representation that constitutes core knowledge of objects has a long evolutionary history, extending deep into human beings' primate heritage.

In this chapter we have seen two distinct ways in which knowledge acquisition concerning objects must transcend core knowledge. First, infants create representations of object kinds that support kind-based processes of individuation and identity tracing that are distinct from those in the mid-level object tracking system. Second, knowledge encapsulated within the object-tracking system somehow becomes explicit. The chapters to come explore additional systems of core knowledge, additional ways in which knowledge acquisition transcends core knowledge, and begins to characterize the mechanisms through which representational systems that transcend core knowledge are constructed.