Chapter 2. Innate Perceptual Representations: the Empiricist Picture

The British Empiricists’ picture of conceptual development finds articulate and ardent defenders to this day. This staying power has two explanations. First of all, the empiricists staked out an ambitious set of phenomena that a theory of concepts must be responsible for. They sought to explain how concepts refer, how people categorize, how concepts function in thought, how human knowledge is warranted, and how human knowledge is acquired. And they offered a theory that accounted for all of these phenomena in a comprehensive, integrated, and unmatched manner. Second, the theory contains important grains of truth.

1. The Empiricist Picture

According to British empiricists such as Locke (1690/1975), human concepts are grounded in a set of primitive ideas. These ideas are *definitional* primitives, all concepts are either primitive or complex, and all complex concepts are defined in terms of primitive ones. The primitive ideas are *innate*; the acquisition of concepts is explained by a specification of the set of innate primitives and by the associative mechanisms through which complex concepts are built from them. Finally, these ideas are *computational* primitives; they are the ideas intrinsically interpreted, such that understanding complex concepts consists of unpacking them into the primitives that define them. And finally, these primitives are *perceptual*. They are the output of sense organs.

That the primitive ideas are perceptual was important to the empiricist program. The empiricist explanation of reference depended upon the view perceptual representations refer to certain aspects of the world by virtue of the operation of the sense organs. That is, they took as an unproblematic foundation for our conceptual life that the operations of our sense organs guarantees that representations of *red* or *round*, for example, are caused by red things or round things. As long as the referential potential of primitive concepts is guaranteed by how perception works, and as long as all concepts may be defined out of primitive concepts, then the referential potential of all concepts is explained. The extensions of complex concepts are determined by their definitional structure and the extensions of the primitives from which they are built. And as long as the primitive ideas were innately interpreted, complex concepts could be understood by translation into the primitives from which they were built. Finally, one concern of the empiricists was epistemological—how knowledge is to be justified. How can I *know* that water is H2O, that I have two feet, or that vixens and foxes? The empiricist view of meaning played at least two important roles in their epistemology. First, it provided a rationale for *a priori* truth. Because the concept of vixen contains the concept fox in its definition, the statement “vixens are foxes” is true in virtue of meaning, is analytic. Thus definitions provide a large class of analytic truths that are immune to refutation; the justification for belief in this class of statements need be sought no further. Second, epistemological considerations were part of the motivation for the empiricists’ identification of primitives with sensory concepts. They attempted to reduce the problem of the justification of contingent statements to the problem of empirically discoverable relations among perceptible properties of things in the world.

This is not the place for a full exposition and critique of the empiricist view of concepts (see Margolis and Laurence, 1999, for a superb collection). Every part of the view has come under fire—from Quine’s and Putnam’s denial of the analytic/synthetic distinction, to Kripke’s and Putnam’s denial of the idea that intensions determine extensions, to Fodor’s critiques of the distinction between primitive and complex concepts. Here I wish to focus on just two points. First, I believe that the empiricists were basically right that there are innate perceptual representations, and that the content of perceptual representations is ensured by evolution. It is because our sense organs work the way they do that humans can see color, movement, shapes, sizes, can taste salt and sweet, can hear tones and noises, etc. Second, chapters 3 and 4 argues that the content of some conceptual representations, those part of core
knowledge, is guaranteed in just this way—through the operation of evolutionarily selected input analyzers. The focus of the present chapter is the empiricist claim that human representational abilities, in the beginning, include only perceptual primitives. Many modern thinkers, even while denying other central tenets of the classic empiricist view of concepts, hold this view. In this chapter I begin to build the argument for core conceptual knowledge by marshalling evidence against it.

2.1. The initial state: perceptual/sensory motor primitives.

Modern adherents to the view that the initial repertoire of mental representations is limited to a set of perceptual or sensori/motor primitives are such major 20th century figures as Piaget (1954) and Quine (1960), as well as many contemporaries who work in the connectionist and the dynamic systems frameworks. Sometimes these writers take an even stronger stand, denying that infants have mental representations at all. Piaget, for instance, considered the sensori-motor stage to be pre-representational; see also Thelan, Schoner, Scheir, and Smith (2000; section 2.10.2 below). Disputes about when infants first have the capacity for mental representation quickly become merely terminological. Perceptual representations are representations, after all, and even reflexes require perceptual representations as input. After the designation of the term “representation” has been settled, the disputes become empirical. Piaget believed memory was one benchmark of true mental representations, which is why he privileged delayed imitation, among other behaviors, as providing unambiguous evidence for them. He believed that there was no empirical evidence for memory in infancy, whereas we now know that even neonates are capable of delayed imitation (e.g., Meltzoff and Moore, 1977), and that the memory capacities of young infants are prodigious (e.g., Rovee-Collier, date). Indeed, in other writings Piaget granted young infants sensori-motor representations. Piaget’s position, then, was that the initial representations of infants are perceptual and motor, serving only sensori-motor reflexes and learned sensori-motor schemata.

Quine was no empiricist. He denied that theoretical terms or the terms in natural languages could be defined in terms of perceptual primitives (or even that the notion of definition made sense; he denied the analytic/synthetic distinction). Nonetheless, Quine agreed that the infant’s initial representational resources were limited to an innate perceptual vocabulary, an innate perceptual similarity space, that he called “a prelinguistic quality space.” In a series of influential writings (Quine 1960, 1970, 1977), he developed three interrelated theses about conceptual development:

1. Infants’ representations are radically different from those of their elders, and are formulated over a perceptual similarity space.
2. The concepts that articulate commonsense ontological commitments are a cultural construction.
3. In the course of mastering natural language, each child acquires adult ontological commitments through a bootstrapping process.

Chapters 5 and 6 will explore the second and third of Quine’s three theses; this chapter concerns mainly the first.

Before we can begin to address the question of whether the empiricists, Quine, Piaget, Thelan, Smith, and others are right that the infant’s initial representational repertoire is limited to perceptual representations, we must agree on what perceptual representations are. We may approach this in two complementary ways: providing a positive characterization of perceptual representations and by contrasting them with representations that are not perceptual. Neither Quine nor Piaget fully characterize the initial representations they attribute the baby, but I believe that they both meant them to be the sense data of the logical positivists, observational properties of entities in the world. Perceptual representations include sensory qualities like color, pitch, loudness, roughness, heaviness, as well as descriptions of the visual array stated in
spatio-temporal vocabulary (shape, properties of motion, and so on). On this view, perceptual representations may be learned—the prototype of a Posner and Keele (date) dot pattern is abstracted from exposure to exemplars each of which is a distortion of the prototype; the representations entering into this learning process, as well as the outcome of the process, are perceptual representations, because they may be completely characterized in spatial terms. Piaget’s sensori-motor primitives are the same; they support current action through reflexes and acquired sensori-motor schemes. Similarly, the representations that articulate Thelen et al.’s motor workspaces are representations of locations or of motor plans; they support current action. Although Piaget thought that these representations existed only when activated by currently perceived stimuli, and thus did not implicate memory except via the accretion of sensori-motor schemata during previous deployment of activity, this is not a necessary part of the view. Memory is an essential part of Thelen et al.’s models, and Quine also believed that children could form associations between entities represented in perceptual similarity space, a form of learning that clearly requires memory. What characterizes these representations as perceptual is their content—they pick out perceptible features of the infant’s world, features that may be expressed entirely in spatiotemporal and sensory vocabulary. Also, insofar as content is diagnosed by examining inferential role, it is important that putative perceptual representations have no intrinsic conceptual roles.

2.2. What are not perceptual representations

Most representations underlying human natural language are not perceptual representations. Human beings represent non-observable entities (beliefs, vital forces, souls, protons), non-observable properties of observable entities (functions, essences), abstract entities (numbers, logical operators), and fictional entities (Gods, ghosts, Hamlet). Such concepts are not themselves output of sense organs. Of course, the empiricist position is that such concepts, although not directly perceivable, are definable in terms of perceptual primitives. Apart from the problem that no adequate definition has ever been provided for any concept (see Fodor, 1981; Laurence and Margolis, 1999), certainly no definitions for the above types of concepts have ever been formulated over perceptual primitives. Laurence and Margolis (1999) quote Locke (1690/1975) characterizing the reduction of the concept lie in terms of perceptual primitives:

1. Articulate Sounds. 2. Certain Ideas in the Mind of the Speaker. 3. Those words the signs of those Ideas. 4. Those signs put together by affirmation or negation, otherwise that the Ideas they stand for, are in the mind of the Speaker.

I think I need not go any farther in the Analysis of that complex Idea, we call a Lye. What I have said is enough to shew, that it is made up of simple Ideas: And it could not but be an offensive tediousness to my Reader, to trouble him with more minute enumeration of every particular simple Idea, that goes into this complex one; which, from what has been said, he cannot but be able to make out to himself.

Laurence and Margolis comment that not only is it not obvious how to continue this decomposition down to sensory primitives, Locke hasn’t even begun the decomposition. Are the concepts speaker, affirmation, negation, standing for any closer to the sensory level than is the concept lie?

This line of argument defeats the position that all human representations are either perceptual, or enriched perceptual representations. However, it does not defeat the argument that the innate primitives are perceptual, as long as one provides a learning mechanism, such as
Quine’s bootstrapping mechanisms, that can account for the creation during development of non-perceptual representations given a beginning stage containing only them. I believe that bootstrapping processes underlie discontinuities in conceptual development (Chapters 7-9), and thus I do not accept a general learnability argument for the impossibility that the initial state may consist only of perceptual, sensory-motor, representations.

The only avenue open to evaluating the empiricist position that starting point of conceptual development consists of perceptual and sensory motor representations alone is empirical. The next several chapters present evidence for a variety of non-perceptual representations in infancy. As I believe that bootstrapping processes are not available to infants (requiring as they do, explicit, external, representational systems such as language and notational systems), then conceptual representations in infancy must either be innate or created from perceptual representations by some process not yet specified. I shall be content if I convince you that infants have representations with conceptual content. To do so, I must characterize particular conceptual representations infants have.

2.4 That some entity is perceived does not make its representation a perceptual representation

Before examining empirically whether infants have at least some non-perceptual representations, more work is needed contrasting perceptual representations with representations that are not perceptual. The commonsense contrast appealed to above is between representations that are the output of sensory transducers or perceptual input analyzer (e.g., pitches and loudness of tones, phonetic features, tastes, aspects of the visual array such as depth relations, figure/ground processes, shape representations, motion representations) and those that are not. This distinction holds even in the face of the obvious fact that we have perceptual representations like dog shaped or Coke taste, representations of what particular kinds look like, sound like, feel like, and so on. That is, we can recognize a chair or a dog upon seeing one, based on perceptual input analyzers (learned, in this case, but that is irrelevant to the point).

However, that we can recognize instances of a given concept on the basis of perceptual input does mean that that concept is a perceptual concept. Take table and dog. Most analyses of artifact concepts like table agree that original intended function is part of what makes an artifact the kind it is, that this fact plays a central role in artifact concepts, and that representations of intended function are not perceptual. And most analyses of natural kinds like dog agree that a hidden essence is part of what makes an individual the kind it is, that this fact plays a central role in natural kind concepts, that essences are in the world and not the mind, and that representations of hidden essences are not perceptual.

Thus, the fact that we have representations of what entities in the world look like, sound like, feel like, and taste like does not mean that our representations of those entities are perceptual. This point becomes important when we consider non-perceptual representations such as contact causality for which there may be innate input analyzers. I have often had the following exchange:

   Me: Representations of contact causality are non-perceptual.
   Other: No they are perceptual. Michotte showed that we can perceive causality in launching events.

But notice, even if the perceptual input analyzers are innate, this argument has no more force than:
Me: Representations of *dog* are non-perceptual.
Other: No they are perceptual. Many people have shown that we can recognize dogs from perceptual input.

*Contact causality* is not a perceptual representation because, as Hume (ref) showed, the attribution of causality goes beyond empirical necessity and sufficiency (perfect co-variation) and also goes beyond the spatiotemporal conditions for perceiving causality in Michotte-like situations. Consider a launching situation: Object 1 approaches Object 2, comes into contact with Object 2, and Object 2 immediately moves off. When some spatiotemporal relations between Object 1 and Object 2 obtain (there must be contact, Object 2’s motion must begin simultaneously with the contact, and Object 2’s direction and velocity of motion must be consistent with Object 1’s and with the angle of contact), people see the motion of Object 1 causing the motion of Object 2. They see Object 1’s hitting Object 2 as launching it into motion.

Under a wide variety of parameters of motion and viewing conditions, the above described spatiotemporal relations between the motion of Object 1 and the motion of Object 2 are necessary and sufficient for the perception of causality (Michotte, 1963). Still, causality (in the world) goes beyond those spatiotemporal parameters of motion and contact, and in the mind as well, the representation of causality goes beyond those spatiotemporal descriptors. In the real world, the Michotte spatiotemporal relations are not sufficient for there being a causal relation between the motion of the two entities (in the Michotte situation itself, where A and B are 2-D disks on a computer screen, the motion of A does not actually cause the motion of B). It is a matter for physics to explain why launching occurs. A and B must be 3-D material bodies before Newtonian mechanics applies. As far as the representation of causality is concerned; the mind supplies the causal attribution in Michotte situations; the spatiotemporal properties are the input to the analyzer and the output is the representation of causality. This representation is not itself perceptual because its content (*cause*) cannot be expressed in spatiotemporal terms. Also, unlike perceptual representations, *cause* has a rich inferential role. Minimally, causal interactions between two events assign very different status to the first and second event (one is the agent and the other the recipient), contact causality is necessary in the explanation of inanimate objects changing state from rest to motion, but not for animate objects, and ultimately contact causality is understood in terms of mechanisms at a molecular level of description.

2.5. Do infants represent the causality in Michotte launching events?

Just when infants interpret Michotte-type launching events in causal terms is debated. The question must be approached in two steps. First, we will want to know whether infants are sensitive to the spatiotemporal parameters of launching. Then, we will want evidence that infants’ representations of launching have at least some causal conceptual role. To anticipate the current state of the art: there is no doubt that infants are sensitive to the spatiotemporal conditions of launching, and although it is much more difficult to shows that they attribute causality in these events, I am convinced that they do.

The literature I review in these early chapters draws upon patterns of looking to diagnose infants’ representations of the world, especially experiments using the violation of expectancy looking time methodology. In this paradigm, infants watch as events unfold before them. On some occasions, a magic trick is performed, creating an impossible or unexpected event. The robust result is that infants look longer at unexpected/impossible events than at ordinary ones, presumably because violations of expectancy are attention grabbing. Babies cannot react to a violation of the expected unless there is some mismatch between their representation of a current
outcome and their representation of the antecedent events, and thus the researcher can use patterns of elevated vs. non-elevated looking times as a source of data concerning the infant’s representations of the ongoing events. In Chapters 3 I shall consider further the nature of the infant’s representations of these events; here we need only indicate their existence and minimal conclusions concerning their content—their extensions and conceptual roles.

The very first experiment using the violation of expectancy looking time method addressed infants’ understanding of contact causality (Ball, 1973). Ball habituated infants to events as in Figure 2.1. An event began with a screen visible on a stage floor with an object, A, partially visible at its right edge. A second object, B, rolled onto the stage from the left, and went behind the screen. After timing consistent with a launching event, object A went into motion and stopped, visible, to the right of the screen. Infants’ looking at the outcome of this event was monitored, and the trial was terminated when they looked away from the outcome array for 2 seconds or more. The whole stage was then covered by curtains, after which the curtains were opened showing the beginning array and the event repeated as before. Infants were shown this event over and over until they habituated to it, that is, until their looking at the outcome arrays was a half of its initial levels.

How had the infants interpreted this event? Did they think that object B had caused the motion of object A via contact? To address this question, Ball then showed infants two new events, in alternation, an expected event (as defined by adult expectations) and an unexpected event. The events were exactly like the habituation event, except that there was no screen. In the expected event, in which contact causality explains the motion of object A, object B emerged from the right edge of the stage, rolled toward and hit object A, after which A went into motion and then stopped as before (Figure 2.1a). The unexpected event was identical, except that object B stopped short of object A, and after a short period of time, object A went into motion and then stopped as before (Figure 2.1b).

The participants in Ball’s experiments (ages 6 to 26 months), looked longer at the unexpected event than at the expected event. Ball concluded that the event in which there was
no contact between the balls was unexpected for infants, just as it is for adults, and thus that infants represent Michotte contact causality. Ball’s experiment was never published, and subsequent studies using the violation of expectancy looking time method use much narrower age ranges of participants. It has, however, been replicated with 6 month-olds (Spelke, Phillips, and Woodward, 1995).

Ball’s experiment succeeds in demonstrating young infants’ sensitivity to the spatiotemporal parameters that determine Michotte launching events, but it falls short of satisfying the second requirement—that they represented the event as a causal one. They may have merely learned a perceptual pattern, like a Posner dot pattern, to the effect that contact of a moving object with a stationary one is associated with simultaneous motion of the previously stationary object.

To show that the infants’ representations have some content that goes beyond generalizations stated in perceptual or spatiotemporal vocabulary poses a significant challenge to the ingenuity of researchers. Many have risen to this challenge. One route taken is to show that infants represent restrictions on when the spatiotemporal features of launching are associated with the movement of one object following the movement of another. For example, in the same experiment in which they replicated Ball’s, Spelke and her colleagues ran one group of infants in a condition in which object A and object B were people. The event began with person B half visible behind a screen, and infants were habituated to A coming into view from the left, going behind the screen, whereupon B went into motion and stopped in full view. Infants in the object version of Spelke et al.’s study (A was a large box; B a chair) showed the same pattern of results as did Ball’s babies; they looked longer at the test events in which there was no contact between A and B. But infants in the people version did not differentiate the two events; a person going into motion without having been contacted by another person did not draw more attention than a person going into motion after having been contacted by another moving person. This experiment is important for its implications concerning infants’ differentiation of people from inanimate objects, but for now the question is whether infants impute causality in the inanimate launching condition. The result is suggestive: the distinction the infant is drawing is between entities that can move on their own and interact at distance (people) and entities that cannot (box/chair). If that is the distinction being made, it implicates causal notions.

Another route taken is to explore whether infants represent causally relevant parameters of launching events. For example, Kotovsky and Baillargeon (refs) have shown that infants represent such generalizations as, for a constant B, the larger A is, the farther B will move. This result is consistent with a generalization stated in causal terms—the larger the object is that hits a another, the farther it will cause that other object to go.

The third route taken is to explore whether infants’ representations give different status to the agent and the recipient in launching events, in the face of failure to do so when two events follow regularly, one after another, but are not causally related. Leslie and Keeble (ref) carried out the following extraordinary study. Two groups of infants were habituated to events in which one object A approached a stationary object B from the left, A stopped, after which B went into motion to the right and then stopped (Figure 2.2). The only difference was that in one group, A stopped short of B, and the motion of B did not begin simultaneously with the stopping of A, whereas in the other group, the spatiotemporal relations between the motions of the two objects specified contact causality. In spatiotemporal terms, each motion could be described in terms of direction and speed of motion of each object and spatiotemporal relations between the two motions (motion of B follows motion of A, either simultaneously or after a temporal gap). If this
is the vocabulary in which the infants were representing the events, then reversing the motion (having A be stationary at the beginning, having B emerge from the right, move towards A, after which A moves to the left and stops) should be equally novel in the two cases. But if the original launching event is represented as A causes B to move, and the original non-launching event is seen as A moves and then B moves, then it is possible that reversing the launching event, in which the roles of the agent and recipient are reversed, will be more noticeable to the baby than reversing the non-launching event. This is what happened: 6-month-old infants dishabituated to the reversal of the launching event (Figure 2.2a) whereas they did not dishabituate to the reversal of the non-launching event (Figure 2.2b). This experiment suggests that young infants represent launching events in terms of concepts that go beyond spatiotemporal descriptors.

None of these experiments provide absolute knock-down evidence that preverbal infants interpret Michotte launching events causally. The empiricist could claim that the representations that underlie the patterns of looking in these experiments are all formulated in spatiotemporal vocabulary, and the empiricist might be right. Just as infants may have learned the contingent relations among the spatiotemporal features of launching, so do they may have learned restrictions on those contingencies stated over perceptual features of the entities involved (Spelke et al.), and generalizations concerning the relative sizes of objects and relative extents of motions (Kotovsky and Baillargeon). The Leslie and Keeble study is more difficult to reinterpret in this way, but one may imagine an empiricist response: infants bring previous learning of the spatiotemporal contingencies launching events to the representation of the events in Figure 2.2a, but no such previous learning is relevant to the events in Figure 2.2b. Exactly why this would lead to an interest in the reversal of the events only in the former case is not clear however. Also Leslie and Keeble did not find differential rates of habituation to the two classes of events.

Although the empiricist has responses to my interpretations of the above three sets of studies, I doubt them, for three reasons. First, attributing the generalization the motion of A causes B to move to the infant is the most parsimonious account of how the young infant learns the generalizations revealed in the above studies. If Michotte contact causality organizes infants’ representations of launching events, we can make sense of their distinguishing agents and
Figure 2.2

DIRECT LAUNCHING

frames: 12  24  1  24  12

total  73

DIRECT LAUNCHING REVERSED

a.

DELAYED REACTION

frames: 8  24  12  24  8

total  76

DELAYED REACTION REVERSED

b.
recipients, of their learning generalizations of what can be agents and recipients, and of their learning properties of motion contingent on which is which. Notice, in the Kotovsky and Baillargeon study, it is important that it is A that is larger; as B becomes larger, B would move less upon being hit by an object A of a constant size. Second, as we will see in Chapter 4, there is a great deal of additional evidence that infants attribute intrinsic agency to humans and other objects that exhibit evidence of intentionality, which suggests that they distinguish intentional causality from contact causality. Third, learnability considerations favor the view that infants represent Michotte launching events in terms of cause. By age 2, children certainly see the causality in launching events, for they mark it explicitly in language, both paraphrastically “A makes B go” and in causal verbs, “A moves B” (Bowerman, ref). The learnability argument then goes back to Hume. If one cannot define “cause” in terms of spatiotemporal vocabulary—where does the explicit concept come from?

Although I find these three arguments persuasive, I do not consider them conclusive. Language learning could conceivably create a representation of causation via a bootstrapping mechanism, and the generalizations infants display in these experiments could be formulated over spatiotemporal vocabulary that is all that is available before that bootstrapping process. Until somebody has sketched such a bootstrapping process, however, there is no reason to doubt that infants represent contact causality as adults do. If so, contact causality is an example of a conceptual representation preverbal infants are capable of forming, a conceptual representation that is the output of the perceptual analyzer described by Michotte.

2.6. Why object is not a perceptual representation

The central case study in this chapter is the concept object, in the sense of representations of substantial, 3-D, material bodies that exist independently of the observer. Are Piaget and Quine correct that: 1) object representations non-perceptual? 2) they not available to young infants? 3) they are built from perceptual primitives in the course of development?

Between them, Piaget and Quine offered several distinct reasons to consider object representations to go beyond perceptual primitives. First, Piaget argued that if perceptual representations are the output of modality specific sensory analyzers, then object representations are not perceptual because they are multi-modal. For adults, a visual representation of an object specifies what it will feel like, where it will be if one reaches for it, and so forth. Piaget, along with the British empiricists and with Quine, believed that infants had to learn the cross-modal correspondences among the perceptual representations of object appearances. This was no problem for their theories; indeed, the British empiricists believed that learning those cross-modal correspondences constituted building the complex concept object. Of course, learning contingencies among perceptual representations in different sense modalities does not require non-perceptual vocabulary. But the learning of such contingencies, Piaget thought, was the first step in transcending the initial sensorimotor primitives. Second, and here Piaget and Quine are in agreement, if perceptual representations are limited to what entities look like, feel like, taste like, move like, objects cannot be represented as individuals that persist through time, independently of the observer. Not only would there be no representations of permanent objects, as Piaget stressed, but the infant would not have access to fundamental quantificational devices. The child could not represent a given object as the same one as one seen earlier.

As Quine saw it, the infant’s representation of all entities is most like how we represent stuffs—ah “that’s water,” “that’s red,” “that’s table, that’s Mama.” As Quine so elegantly put it:
We in our maturity have come to look upon the child’s mother as an integral body who, in an irregular closed orbit revisits the child from time to time; and to look upon red in a radically different way, viz., as scattered about. Water, for us, is rather like red, but not quite; things are red, stuff alone is water. But the mother, red, and water are for the infant all of a type; each is just a history of sporadic encounter, a scattered portion of what goes on. His first learning of the three words is uniformly a matter of learning how much of what goes on around him counts as the mother, or as red, or as water. It is not for the child to say in the first case “hello mama again,” in the second case “Hello, another red thing;” and in the third case, “hello, more water.” They are all on a par: “Hello, more mama, more red, more water” (Word and Object, p 92)….

The child can learn “apple” as well, as a mass term for sporadic uncut apple stuff.” (Word and Object, p 92).

The contextual learning of these particles goes on simultaneously, so that they are gradually adjusted to one another and a coherent pattern of usage is evolved matching that of one’s elders. This is a major step in acquiring the conceptual scheme we all know so well. For it is on achieving this step, and only then, that there can be any general talk of objects as such. (Ontological Relativity and Other Essays, p 9-10).

It should be difficult to understand how Quine is imagining the initial state of the baby, for he is imagining a representational system very different from yours and mine. Quine’s view can be schematized as follows. Imagine a portion of bottle experience that we adults would conceptualize as a single bottle. (Figure 2.3a) Babies learn associations between portions of bottle experience and other experience—the taste of milk or juice or the sound of the word “bottle.” Now imagine a portion of bottle experience that we would conceptualize as three bottles. (Figure 2.3b) The infant would also expect to obtain milk or juice or hear the word “bottle” when confronted with this experience. What the infant does not do is represent “a bottle” differently from “some bottles” or “three bottles,” nor is she capable of representing “this bottle,” in the sense of the same one, different from another. One way of stating Quine’s hypothesis, as I construe it, is that babies represent no sortal concepts, no concepts that provide conditions of individuation and numerical identity, no concepts that divide reference.

The alert reader may have noted an inherent contradiction within this picture. It is shape that distinguishes experiences of bottle from experiences of toy car, and shape is a property of individuals. So how can the child learn associations between experiences of bottle and other experiences without representing individual bottles? This contradiction is only apparent. Consider entities we quantify over as mass nouns, such as (in English) spaghetti and macaroni. The shape of individual pieces of each of these (see Figures 2.3c and 2.3d) is important for telling the difference between spaghetti and macaroni, and some representation of those individual pieces must enter into the representation of shape. But our concepts, spaghetti and macaroni (and the words “spaghetti, macaroni”) do not quantify over those individuals. Similarly, we can represent the shape of a scattered portion of sand, arranged, for example, into an S, and when we refer to it as “a portion” or “an S,” we are quantifying over that individual. But when we think of it as sand we are not. Quine’s proposal is that the infant’s representational system only categorizes the world in terms of perceptual primitives that capture similarity of sporadic encounters of bottle, sand, mama, water, red, table. Of course, the child’s perceptual
system must pick out individuals in order to represent shape, to determine what to grasp, and so on. This is part of what Quine (1974) meant when he claimed that the child is inherently “body minded.”

Figure 2.3

Quine’s view of the depth of the difference between a representational system formulated over an innate perceptual quality space and one formulated over concepts with the quantificational structure of natural languages was emphasized when he insisted that commonsense ontology is a cultural construction, just as the concepts that articulate scientific theories are cultural constructions: “Theory may be deliberate, as in a chapter on chemistry, or it may be second nature, as in the immemorial doctrine of ordinary enduring middle-sized objects” (Word and Object, p 11). “Analyze theory-building how we will, we must all start in the middle Our conceptual firsts are middle-sized, middle-distanced objects, and our introduction to them and to everything comes midway in the cultural evolution of the race (Word and Object, p 4-5).”

What Quine is saying is that human beings, through cultural processes, have laboriously constructed the commonsense ontological commitments of natural languages, and that these differ from the representational resources provided by evolution (the innate perceptual similarity space). There are many passages concerning the conceptual consequences of these achievements. For example, “Once the child has mastered the divided reference of general terms, he has mastered the scheme of enduring and recurring physical objects...After this “Mama” in particular gets set up retroactively as the name of a broad but recurrent but withal individual object, and thus a singular term, *par excellence.*” (Word and Object, p 95).
Finally, Quine envisioned a bootstrapping process through which the child mastered the ontology of enduring and recurring physical objects (see quote 3 above). The child bootstraps the new ontology by gradually learning the quantificational devices of natural languages—quantifiers, determiners, the *is* of numerical identity, and so forth. Chapters 7-9 attempt a sympathetic characterization of bootstrapping processes, arguing for their role in the construction of new representational resources.

What Piaget wrote about under the rubric of “object permanence” comes to the same thing as what Quine wrote about under the rubric of “divided reference” and quantificational capacities. When babies reach for a hidden object, and we attribute to them an appreciation of object permanence, we assume that they represent the object they seek as the *same object, the same one*, that they saw disappear. Otherwise, it’s not object permanence, but is rather some learned contingency such as “reach where I saw some visual property disappear and some visual property or tactual property will be there.” The latter generalization is formulated in the language of perceptual primitives; whereas “reach for the object that went behind the screen; it will still be there” is not, for “the object” and “it” pick out a single individual persisting through occlusion. The criteria for individuation and numerical identity for ordinary objects go beyond perceptual primitives. In the adult state representations of objects are constrained by the principle of spatiotemporal continuity (objects do not go into and out of existence) and solidity (one object cannot pass through the space occupied by another). Although perceptual primitives can specify a currently perceived, bounded, entity, and its current path of motion, they do not specify that the entity continues to exist when we lose perceptual contact with it. This construal is provided by the mind, and the question raised by both Piaget and Quine was how representations of permanent individuated objects, quantified as discrete individuals tracked through time, come to be formed.

Before we consider how infants form representations of object permanence, how they create criteria for individuation and numerical identity of objects, we must consider when they do so, for Piaget’s and Quine’s theories of how depend crucially upon when. So too for cross-modal representations of objects. Piaget claimed that cross-modal correspondences among perceptual properties of objects were learned by 7 months (by the end of the stage of secondary circular reactions), but that infants’ representations of objects as permanent, existing apart from their own sensori-motor schemas, emerged only between 18 and 24 months (heralding the end of the stage of sensori-motor intelligence). Quine claimed that the capacity to represent “objects as such” emerged only when the child mastered the quantification devices of natural language (between ages 2:0 and 3:0, e.g., see Gordon, date, Soja, 1992; Soja, Carey and Spelke, 1991). Thus, demonstrations that 2- to 6-month-olds have these representational capacities defeats Quine’s and Piaget’s proposals. Nonetheless, the question of innateness is still open, for infants might form these representational capacities from perceptual primitives during the first 2 months of life. At the end of this chapter, I shall return to the question of innateness.

2.7. Piaget’s and Quine’s evidence.

Quine, a philosopher, did not consider actual empirical evidence for his claim that the initial state consists solely of perceptual representations. Rather, he discussed possible observations, considering whether they could possibly show that prelinguistic infants’ representational capacities are the same as yours or mine. He argued that any piece of behavior we observe is consistent with radically different ontological commitments. The child who points to a bottle and says “bottle,” or who picks up a bottle and drinks from it, *may* have the capacity to represent to individual bottles and to represent generalizations such as “that bottle has milk in it,” or *may*
have simply learned associations between perceptual features of bottle, on the one hand, and a spoken word or an action on the other. Of course, it is this line of argument writ large that ends in Quine’s views of radical indeterminacy, for the same considerations bear on adult linguistic capacities as well. On Quine’s view, ontological commitments are fixed only up to the indefinite number of schemes that are consistent with the grammatical commitments of a given language. I believe Quine is wrong, and we can bring evidence to bear on the child’s quantificational capacities and ontological commitments.

Piaget’s genius was at bringing empirical data to bear on age-old philosophical questions, and his experiments on object permanence are justly among the most celebrated in developmental psychology. He reported observations that are consistent with the claim that young infants lack representations of permanent, multi-modally specified, objects. With respect to intermodal correspondences, he made observations of infant’s being startled when they made a fast movement of their own hand across their visual field, and assumed this meant that they did not know what their own hand looked like, nor could they related a perceptual representation of a location of a visually located visual experience with a proprioceptive representation of the location of limb. He also made observations of infants examining their own hands or feet and assumed that these provided the experience the infants needed to make intermodal representations of their own bodies. These could then scaffold associating the visual, tactual, and spatial correspondences among the perceptual representations of external objects.

With respect to object permanence, Piaget made a two-part empirical argument that infants did not represent objects as spatio-temporally continuous. First, he showed that below 8-months of age or so, an infant reaching for a desired object will abort the reach if the object is hidden under a cloth or cup or if it is hidden behind a screen, in spite of the fact that the infant has the motor capacity to remove the barrier. Piaget thought that this behavior showed that the infants did not represent the object as continuing to exist when out of sight. Second, he argued, (like Quine) that the 8-month-olds’ success does not necessarily mean that they do represent objects as existing spatio-temporally continuously. Rather, he took the A-not B error to show that they do not do so. After retrieving an object in hidden location A, if it is next hidden in location B, the infant will search again in location A. Piaget’s interpretation was that infants had simply learned a rule, “look where something has disappeared and something interesting will happen,” rather than they were tracing the identity of the object through changes in location. It was not until infants could solve the hidden displacement problems (18-months or so), that Piaget was willing to credit them with a representation of object permanence.

Although these Piagetian observations are extremely reliable, having been replicated countless numbers of times, methodological advances have provided a wealth of data that Piaget underestimated the representational capacities of young infants.

2.8. Intermodal representations

The empiricists believed that learning the intermodal correspondences between perceptible properties of objects (how visual appearance of texture is correlated with how that texture feels when touched, how visual appearance of shaped is correlated with how that shape feels when touched, etc.), as well as learning the intermodal correspondences between visually specified locations and the effects of reaching to proprioceptively specified locations, is the process through which representations of objects are built. For the empiricists, there was nothing more to object representations. Piaget also believed that such intermodal representations had to be learned, and considered this learning an essential part of the process through which non-perceptual representations of objects are built.
The empiricist position misses the mark in two ways. First, even once all those intermodal representations are formed, infants still would not have conceptual representations of individuated, spatiotemporally continuous objects that exist independently of themselves. Second, there is now massive evidence that skeletal intermodal representations are innate, and certainly not learned through the associative mechanisms Piaget and the empiricists imagined (string of references). Neonates orient visually to a location specified by a sound, prereaching infants attempt to grasp only objects that are within reach, and so on. Two experimental results can stand as examples from this large and convincing literature.

Meltzoff and Moore (ref) allowed neonates to suck on a strangely shaped pacifier—either a smooth cube, or a sphere with bumps all over it. The babies were not allowed to see the pacifier. At the same time (or later in some experiments), the infants were shown two pictures—one of a cube and the other of a sphere with bumps. The babies preferentially attended to the picture that matched the pacifier on which they sucked. Thus, the infants innately recognized the correspondence between the visually and tactually specified shapes/textures.

A second example also comes from Meltzoff and Moore’s laboratory. They showed that neonates would imitate the facial gestures of an experimenter (mouth opening, tongue protrusion.) Chapter 4 considers the significance of this result for the characterization of core knowledge of conspecifics, but for now it is enough to show innate knowledge of the correspondence between what another’s face looks like and the actions and feel of their own face.

That representations of people and objects, including their locations in space, are specified intermodally in neonates is what would be expected on the hypothesis that what is being represented are 3-D objects (including people) that exist independently from the child and the child’s actions on them.

2.9 Criteria for individuation and numerical identity of objects; object permanance

By 4-months of age, infants represent objects as spatiotemporally continuous. Not only do they represent objects as continuing to exist behind barriers, they also take evidence of spatio-temporal discontinuity as evidence for numerical distinctness. Many violation of expectancy looking time studies support these generalizations; I begin with just two.

Figure 2.4a schematically depicts an event shown to infants (4 ½-month-olds in Spelke, Kestenbaum and Wein, 1995; 10-month-olds in Xu and Carey, 1996). Two screens are placed on an empty stage, and objects are brought out, in alternation, from the opposite sides of the screens and then returned behind them. Two objects are never simultaneously visible, and no object ever appears in the gap between the two screens. In some studies, infants are fully habituated to these events; in other studies, they merely are familiarized to them by showing some number of iterations. The question is, how many objects are involved in this event? For adults, the answer is unambiguous: at least two. This event cannot consist of a single object going back and forth because its path would be spatiotemporally discontinuous; it would have to dematerialize behind the right hand screen and rematerialize behind the left hand screen.

We ask infants how they represent the events by removing the screens and showing them one of two outcomes: the expected outcome of two objects or the impossible outcome of just one object (thanks to a magic trick; one of the objects is surreptitiously removed through a trap door in the rear of the stage). Typically, in these studies, after familiarization or habituation to the events, infants are shown a test outcome in which the screen is removed revealing either one object (the impossible outcome) or two objects (the possible outcomes). Then the stage is
cleared, the familiarization event repeated and the other test outcome revealed. Usually there are 3 pairs of possible/impossible trials, alternating, with order counterbalanced across infants.

In these studies, infants look reliably longer at the impossible outcome of two objects than at the expected outcome of one (see Figure 2.5a from Xu and Carey, 1996). Of course, one must worry about alternative explanations for any given pattern of results. In this case, perhaps infants are not representing the path of the object(s) emerging from behind the screen at all. Perhaps the most salient aspect of these arrays during the familiarization part of the experiment is that there are two screens. The preference for one object in the outcome arrays may be a **novelty** preference—an array of one object is more novel, relative to the two-screen familiarization arrays, than is an array of two objects. This alternative requires that infants distinguish one from two, but not that they represent the objects continuing to exist behind the screen nor that they use evidence regarding spatiotemporal continuity as a basis for computations concerning numerical identity.

A control for this alternative is to show the object appearing in the gap during familiarization (Figure 2.4b). The simplest interpretation of this event is that it involves a single object going back and forth behind the screens, and indeed, that is the interpretation 10-month-old infants make. When the screen is removed and the outcomes revealed, infants now look longer at the two object outcomes than at the one object outcomes (Figure 2.5b, Xu and Carey, 1996). The differentiation of patterns of looking in the discontinuous event and the continuous event shows that infants indeed analyzed the paths of the object(s) emerging from behind the screens and established representations of two objects in the two object events on the basis of spatiotemporal discontinuity.

---

**Figures 2.4a, 2.4b**

---

In Spelke et al.'s (1995) original results, 4 ½-month-old infants showed no preference in the continuous motion condition, which, in fact, is indeterminate with respect to the number of objects in the array. What is important is the infants' differentiation of the continuous and discontinuous events.
Chapter 2

Figure 2.5

The Spelke et al. experiment involved 4-month-old infants. Aguiar and Baillargeon (date, date), using a slightly different method, have shown that 2-month-old infants also expect objects to move on spatiotemporally continuous paths, even through occlusion, although it is not until 3-months of age that infants can use evidence of spatiotemporal discontinuity to posit a second object in such events.

Wynn’s (1992a) famous “addition/subtraction” experiments support the same conclusions. Wynn used the violation of expectancy looking time paradigm to explore whether infants could update a representation of a hidden object or objects when additional objects were added or subtracted from the set. The first study tested 4½-month-olds on $1 + 1 = 2$ or $1, 2 - 1 = 2$ or 1, and $1 + 1 = 2$ or 3 events. Take $1 + 1 = 2$ or 1 as an example. The familiarization events were as in the top panel of Figure 2.6. Infants watched as a single object was placed on an empty stage, and a screen was rotated up that hid it. Then the infants watched as a hand brought in a second object and was withdrawn empty. The screen was then lowered, revealing either the expected outcome of 2 objects, or the unexpected outcome of 1 object. Looking times to outcomes of 1 and 2 objects in this condition were contrasted with those from the $2 - 1 = 2$ or 1 condition (in which case, the 2-object outcome is unexpected and the 1-object outcome is expected). Infants’ patterns of looking were different in the two conditions; in the subtraction condition they looked reliably longer at the 2-object outcome whereas in the addition condition they did not.

Wynn’s results have been replicated in many laboratories. Koechlin, Dehaene and Mehler (1996) found the exact same pattern of data with 4-month-olds, as did Feigenson, Carey and
Spelke (1992a) with 7 month-olds. The Koechlin et al. study is particularly interesting because the objects behind the screen were on a rotating plate, such that the infants could not predict the spatial layout of the outcome arrays. Outright success in the $1 + 1 = 2$ or $1$ events on their own was obtained by Simon, Hespos and Rochat (1995) at 4½-months of age, and by Uller, Huntley-Fenner, Carey and Klatt (1999) at 8-months of age. Wynn also found infants succeeded in the $1 + 1 = 2$ or $3$ condition, looking longer at the unexpected outcome of 3 objects. Infants’ attention is drawn when any number of objects other than precisely two is revealed after a $1 + 1$ event.

The implications of these results for our understanding of infants’ representation of number will be explored in Chapter 7; here I wish to emphasize their implications for the Quinian/Piagetian position. To succeed on Wynn’s tasks, infants must represent the object as continuing to exist behind the screen. Furthermore, because the objects are physically identical, the child must use spatio-temporal evidence as a basis for individuation. The infant has no other information relevant to whether the second object is numerically distinct from the first. In the $1 + 1$ event, the infant must represent the object behind the screen, use the fact that the object being introduced in the hand is spatio-temporally distinct to represent it as a numerically distinct object, and update the representation of the hidden array by including a representation of a second hidden object. Not only do these experiments reveal that infants expect objects to be spatio-temporally continuous, they also show that infants’ object representations are governed by criteria for individuation and numerical identity. Contrary to Quine, infants command the logic of divided reference; they distinguish one object seen on different occasions from two numerically distinct objects.
2.10. Why do young infants fail on search tasks?

By 4 months of age, at least, infants represent objects as spatiotemporally continuous, tracking individual objects through space and time, even when occluded (see below for relevant evidence from 2-month olds). Why then, do they fail in search tasks (Piaget, date) or in other tasks that require means-ends planning (Munakata, McClelland, Johnson, & Siegler, date)? Although the answer is not settled, various distinct, but not mutually exclusive, possibilities have been offered. All rely on the observation that the capacity to represent some aspect of the world is not an all or none matter. Representations are graded in robustness or strength (Munakata et al., date; Uller et al., 1999), are constructed in real time, and subject to multiple interacting influences during the processes of construction (Thelen et al., 2000). Furthermore, there are many different visual and motor maps of the world in the nervous system, and it is possible that the representations that play a role in guiding different actions (such as reaching and looking) differ in some respects. For example, Munakata et al. suggested that it is possible that more robust representations are required to support reaching than to evaluate consistency of visual models of the world. Others have noted that the Piagetian tasks differ from the looking time studies in requiring means-ends planning and various executive functions supported by the frontal cortex (maintaining a representation in short term memory, inhibiting competing responses). These processes have a developmental course that is partially independent of the capacity to represent objects.
For the sake of illustration, let’s see how some of these ideas play out in understanding the A-not B error in basic object permanence tasks, which wanes between ages 8 and 12-months or so.

2.10.1 Frontal cortex maturation

Diamond and Goldman-Rakic (1989) began with the observation that the A/not B task closely resembles a task used to diagnose frontal function in monkeys; delayed response (DR). In DR, an item (usually food) is hidden in one of two wells, a delay is imposed in which the animal is not allowed to orient toward the correct well, and the animal is then allowed to search for the item. As in the A/not B task, a crucial determinate of success in DR is whether the search in immediately previous trials is to the same well or to a different well than the current trial.

There is conclusive evidence for frontal involvement in DR (see Diamond, 1991, for a review). Lesions in prefrontal cortex (specifically dorsolateral prefrontal cortex) of adult monkeys disrupt performance. Monkeys with such lesions can still succeed at the task when there is no delay, but performance falls apart at delays as short as 2 s. Lesions in other memory or visual systems (such as the hippocampus or parieto-temporal areas) do not affect DR. Also, there is excellent evidence for a maturational contribution to the development of DR during infancy. In rhesus monkeys, 1.5-month-old infants perform on DR as do adults with lesions in the dorsolateral prefrontal regions. Between this age and 4 months of age, the delay that can be tolerated increased from 2 s to 10 s or more; 4-month-old infant rhesus monkeys perform as well as do adults with intact prefrontal cortex. That maturational changes in prefrontal cortex play some role in this improvement is shown by the fact that lesions in this area at 1.5 month preclude the developmental improvement in DR, and the same lesions at 4 months have the same effect on performance on DR as do such lesions in adulthood—to wit, disrupt it to the level of 1.5-month-old infants.

Diamond (1991) has amassed considerable evidence that the maturational change in prefrontal dorsolateral cortex taking place in infant rhesus monkeys between ages 1.5-4 months occur in infant humans between ages 7.5 and 11 months and at least partially underlie the developmental changes seen in Piaget’s Stage IV of the object concept. Diamond gave the same version of the A/not B task to human infants at this age, to infant rhesus monkeys, and to adult rhesus monkeys who had been lesioned in the prefrontal dorsolateral cortex. She found that the developmental changes in human infants matched, in parametric detail, those of the monkeys, except that the development was a bit slower in humans (over 2.5 months in monkeys, over 4.5 months in humans). In both species, the delay at which the A/not B task was solved increased from 2 s at the youngest age to 10 s or more at the oldest age. In both species, errors were predominantly on trials in which the correct choice differed from the correct choice on the previous trial (i.e., switch trials). In both species, details of the infants’ behavior on the switch trials suggest they represented where the objects was; sometimes the did not even look in the well they had uncovered before reaching for the correct well, and sometimes they stared at the correct well even as they reached for the incorrect one. These behaviors occurred at comparable rates in the two species. Finally, the adult rhesus monkeys with lesions in the prefrontal dorsolateral areas, as expected, failed the A/not B task at delays over 2 s (like the 1.5-month-old rhesus and the 7.5-month-old humans), and made errors predominantly in the crucial switch trials in which the bait was placed in a different well from that of an immediate preceding successful trial.

Diamond concluded that immaturity of dorsolateral prefrontal cortex contributes to the 7.5-month-old’s failure on the A/not B task. Seeking convergent evidence for this conclusion,
Chapter 2

Carey

she reasoned that if maturation of the structure underlies the parametric improvement on this task over the next 5 months or so, then other tasks that diagnose prefrontal dorsolateral function in primates should show a parallel course of development. She confirmed this prediction in a series of elegant studies of babies reaching for objects in transparent Plexiglas boxes. Problems of differential difficulty are posed for the infant as a function of where the opening of the box is placed. Young infants (7.5-month-old humans, 1.5-month-old rhesus) cannot solve this problem unless the direct line of sight between the infant and object is through an opening. If the opening is to the side, for example, infants of both species of these ages keep reaching directly for the object, hitting the Plexiglas wall, and trying again and again until giving up in frustration. Diamond charted a series of stages infants between 7.5- and 12-months go through before complete success at this task and showed that infant rhesus monkeys go through parallel stages between ages 1.5 months and 4 months, and that adult rhesus monkeys with lesions in prefrontal dorsolateral cortex fail at this task, performing like 1.5-month-old infants of their species.

There is no obvious conceptual similarity between the A/not B task and transparent box task. In the former, the object is hidden, and memory is a critical component (performance is a function of delay). In the latter, the object is visible through the box, so memory plays no role whatsoever. What unifies these two tasks is their reliance on intact, functioning, dorsolateral prefrontal cortex. Functionally, it is likely that the aspect of executive function being tapped in both tasks involves inhibiting a prepotent response (reaching along the direct line of sight in the transparent box task, repeating the previously successful reach in the A/not B task). Also, prefrontal cortex is crucially involved in working memory, a critical component of the A/not B task. Diamond argues that these are aspects of executive function supported by the prefrontal cortex, and these are not required in the violation of expectancy looking time studies. On Diamond’s story, then, the A/not B error does not reflect a limit in the infant’s representation of objects as spatiotemporally continuous, continuing to exist when occluded, but rather reflects immature executive function that limit the means/end problem solving of infants under 1 year of age.

2.10.2 The dynamic systems account

In a series of influential writings Thelan, L. Smith, and their colleagues (e.g., Thelen et al., 2000) have discovered several new phenomena and systematized the empirical literature concerning the A/-B error. They argue that the error could arise from complex interactions among the multifaceted processes that enter into motor planning, processes that unfold over time. Thelan, Smith and their colleagues stress that whether the infant makes the error or not is dependent upon many factors, such as how many repetitions of hiding at A before the switch to B, the delay, the salience of the object, the distinctiveness of the two locations, whether the infant is in the same position during the A trials and the first B trials, and so on. One important finding is that under some circumstances, the likelihood of reaching, and the probability of the error, does not depend upon a hidden object at all; if cued by the experimenter’s waving the lid of the container, the infant will reach for the visible lid and show the A/-B error.

In Thelan et al.’s model of motor planning, three distinct representations are built up over time, each having it’s own dynamics (rate of build up, capacity for stability and self-maintenance, time course of decay), and interact in a common motor workspace to create a plan to reach to A or B (or neither). The three distinct representations are a representation of the task environment (that establishes the locations of A and B, maintains them as distinct, as equally or differentially salient), a representation of the cued location of a given trial, and a representation of the previous movements, especially the immediately preceding one, in which this
representation is influenced by the entire history of movements. These representations are integrated in the process of planning a movement; a movement to A or to B ensues when a threshold is reached. The various context effects discovered and reviewed by Thelen and Smith are modeled in terms of parameters that influence the dynamics of the formation and maintenance of each of the three types of representations in motor space, and they model the developmental change between 8 months (A/-B errors likely at delays greater than a few seconds) and 12 months (A/-B errors unlikely, within a wide range of task parameters, at delays as long as 10 seconds) in terms of a change in a parameter they call “cooperativity,” that reflects the differentiation within motor space and the capacity for creating and maintaining a stable representation of the cued location.

Although Thelen’s and Smith’s account differs from Diamond’s in many respects, both place the A/-B error in the context of the interaction of two different memories—memory for the cued location (or for the object’s location) and memory of the past action (see also Munakata, 1998). Memory of the past action has a much longer time course of decay, the limits of which, as Thelen et al. point out, has not been systematically studied. If the processes that form and maintain the short term memory of the cued location are fragile, an A/-B error is thereby likely to occur. In sum, Thelan and Smith, Diamond, and Munakata all agree that the A/-B error arises from the interaction of object or location representations with other representations involved in the planning of a reach.

Thelen, Smith, and their colleagues draw what seem to be stronger conclusions than those outlined above. They sometimes deny the usefulness of the construct “object representation” or even “representation” at all, and argue that to ask when children “know” that objects continue to exist when behind barriers is a badly mistaken question, because knowledge is always manifest in behavior and thus its expression subject to the dynamic interaction of many different processes. Although this latter point is certainly true, it does not follow that the notion of “representation” is bankrupt. Indeed, their own model explicitly depends upon three different types of representations—of the task context, of the cued location, and of past acts. That these representations are formulated over motoric space, evolve over time, and interact in complex ways, does not make them any less representations.

Although the Thelan et al. model crucially depends upon representations, these representations are certainly sensori-motor ones. Representations of objects, even perceptual ones, play no role in the model at all. The representations are representations of locations, with strengths determined by salience. The only possible role for perceptual representations of the objects is that their salience might affect the degree of activation of the location in which they were hidden. In support of the claim that representations of objects are playing no role in this task, Smith et al. (1999) discovered that infants will reach to the lids even when there are no objects in them, and that merely waving one of the visible lids, or touching it, would induce a reach to a particular lid. That is, a wave or a touch would serve as a specific cue to a location on a particular trial.

It is not surprising that a model of the last stages of planning a reach is formulated in a motor work space that includes representations of locations, but it is also unlikely that a full model of the dynamics of the planning, memory, and motivational processes that interact in determining a reach can dispense with representations of the goal of the reach. Could it be true that representations of specific unseen objects do not ever guide reaches at the ages of children of the age of the A/-B error? I think not.
Van de Walle, Carey, and Prevor (2000) developed a violation of expectancy search task that yields data showing that representations of objects guide the reaches of 12-month-old infants. Infants are introduced to a box into which they can reach but cannot see. One object (e.g., a toy shoe) is removed from the box, shown to the baby, and replaced, twice. Or two objects (a toy shoe and a toy duck) are removed from the box one at a time, shown to the baby, and replaced in it. Thus, in both of these conditions, the baby has seen two episodes of reaching into the box. The box is pushed within the baby’s reach. Babies reach into the box to retrieve the objects, and the crucial manipulation in this study is that on two-object trials, one of the objects has been surreptitiously removed before the box is handed to the baby. After the baby removes the first object (e.g., the toy shoe in both 1-object and 2-object trials) the dependent measure is amount of search time for a second object. Success on this task is longer searching on 2-object trials because there should be a second object in the box than on 1-object trials, because the only object the child saw emerging from the box has been retrieved. Twelve-month-olds succeed in this version of the task.

Apparently, infants of this age can represent the difference between one and two objects being in the box, and their reaches into the box are guided by representations of the objects within it. Thus, a full model of the planning process must contain representations of the hidden objects themselves, not only the locations to which the child will reach.

What about the claim that the question of when the baby “knows” that objects continue to exist is misguided? In Chapter 3 I will address the question of in what senses core knowledge is “knowledge,” that is, what kinds of representations are at issue here. But is it a misguided question to ask when in developmental a given representational capacity first becomes available? Thelen et al. argue that it is, because the manifestation of any given representation is so exquisitely context specific. For sure, representations are constructed over time, are graded in strengths, and interact in the control of action. Nonetheless, for any representational capacity we may want to focus on, it is a meaningful scientific question whether there is any context in which it will be constructed, and if not, whether this is because the child does not yet have the competence to do so. This book will present evidence for many such cases (the natural numbers, the concept of a living thing, the concepts of weight and density), and will address how new representational capacities are constructed.

2.1.1 Are object representations innate?

I have argued that very young infants represent objects as spatiotemporally continuous. The computations through which young infants establish representations of objects embody criteria of individuation and numerical identity. Contrary to Quine, a child does not need the ladder constructed from the explicit quantificational devices of natural language in order to create representations of objects that divide reference, that distinguish between the same one and a different one. Contrary to Piaget, a child does not need the full period of sensori-motor development (until 18- to 24-months) to create representations of enduring objects that exist through occlusion. Still, the youngest age of participants in the violation of looking time studies reviewed so far in this chapter so far is 2 months. Is it possible that younger babies’ representations are formulated over a perceptual/motor primitives? Could the capacity to represent and quantify over objects displayed in these experiments be built between birth and 4-months of age?

The question of when infants first represent objects as spatiotemporally continuous is very much open. There is one compelling piece of empirical evidence that suggests that the object representations that articulate 2-month-olds’ (and older infants’) representations of events are not
innate. The phenomenon in question is the capacity for amodal completion of single objects, two ends of which protrude from behind an occluder (see Figure 2.7a). This phenomenon differs from those discussed so far in this chapter, for it does not concern when infants represent whole objects that disappear behind barriers as continuing to exist there. Nonetheless, at issue is the infant’s criteria of object individuation. Under what circumstances, if any, does the infant establish a representation of a single, spatiotemporally continuous, object extending behind the barrier, rather than two numerically distinct objects? Kellman and Spelke (1983) used the violation of expectancy looking time method to answer this question. They found that if the visible ends of the occluder move together, 4-month-old infants establish a representation of a single object, as shown by the fact that upon removal of the barrier, they look longer if a broken rod (Figure 2.7b) is revealed than if a continuous rod (Figure 2.7c) is revealed. In a series of wonderful studies, Johnson and Aslin (refs) have shown that 2- to 4-month old infants are sensitive to almost all of the same information that adults are in computing representations of a single rod in this situation, but that 2-month-olds need more redundant information than do 4-month olds.

Newborns, however, are different. Newborns show the opposite pattern (Slater, Morrison, Somers, Matlock, Brown, and Taylor (1990). Habituated to the array in Figure 2.7a, they look longer at the completed rod (Figure 2.7c) than the broken rod (Figure 2.7b), as if the latter is a
novel stimulus for them. Slater et al.’s findings leave open the possibility that between birth and 2 months of age, infants learn that common motion of two visible portions of objects protruding from behind a barrier are likely to be part of one and the same object. However, it is also possible that maturational events underlie the change between the neonatal representational capacities and those of the 2-month-olds, consistent with a nativist stance. A third possibility is that neonates need more redundant information still, compared to 2-month-olds, just as 2-month-olds do compared to 4-month-olds, and the pattern of looking reveals a familiarity preference rather than a novelty preference. Upon meeting the habituation criterion, the neonates may still be in the process of building the representation of a single object.

It will be difficult to decide, empirically, between these competing interpretations of the neonate pattern of looking. However, other considerations bear on the issue of whether the capacity for amodal completion of objects behind barriers is innate. Those who favor a learning account of the change between birth and 2-months of age need to sketch one. What learning process could create representations of complete objects that persist behind barriers taking only perceptual primitives as input? Similarly, how could infants learn that whole objects that disappear completely behind barriers continue to exist there?

It is easy to imagine how infants might learn such generalizations as:

1) If closed shape1, certain color1, texture1, disappears by continuous deletion along one boundary behind closed shape2, certain color2, texture2, then with certain probability, closed shape3, with color3 and texture3, will emerge by continuous undeletion along same boundary on opposite side of closed shape 2. Shape3, color3, and texture3 will be similar to shape1, color1 and texture1.

2) If closed shape1, certain color1, texture1, disappears by continuous deletion along one boundary behind closed shape2, certain color2, texture2, and if closed shape2 moves away, with certain probability, closed shape3, color3 and texture3 will become visible in location previously behind closed shape2. Shape3, color3 and texture3 will be similar to shape1, color1 and texture1.

3) If closed shape 1, certain color1, texture1, disappears partially by continuous deletion behind closed shape2, such that portions of closedshape3 and closedshape4 remain visible, and such that closedshapes3 and 4 are similar to parts of closedshape1, then, with certain probability, the motions of closedshapes 3 and 4 will be perfectly correlated.

4) If the motion of shape1, color1, texture1 is perfectly correlated with the motion of shape2, color2, texture2, and each is partially behind shape3 (with behind indicated by binocular cues, t-junctions, etc), then with certain probability, if shape3 is removed, closed shape4 will be revealed, and parts of closedshape4 will be similar to closed shape1 and to closed shape2.

Statistical analyses, for example of the sort so well modeled in connectionist architectures, could accomplish this learning, and indeed, there are successful models that do just that (Munakata et al., date) However, these generalizations are not stated over object representations. Furthermore, even if they were, generalizations 1 and 2 would not constitute representations of object permanence, unless the system represents the object as the same one that went behind the barrier. As Marcus (2001) points out, either the current simulations cannot do so (e.g., Munakata, et al., 1997), or they build in this capacity from the beginning (Mareschal, Plunkett, and Harris, 1995), thus accomplishing interesting learning, but not the learning of spatiotemporal continuity itself. Similarly, even if generalizations 3 and 4 were stated over representations of
objects and parts of objects, they would not constitute amodal completion of an object unless they represented the completed object as the same one as unites the parts that had been visible before.

The debates over whether connectionist models could take perceptual input and construct representations of objects that embody criteria of individuation and numerical identity engage the learnability issue in just the right way. Any learning model that could accomplish this feat would defeat the learnability argument that object representations cannot be built from perceptual primitives. It is still an open question whether one can imagine, in principle, a learning mechanism that could accomplish the task. Of course, even if we could imagine one, we wouldn’t know that we were right. It would still be a logical and empirical possibility that object representations are innate in human infants, just as representations of the night sky are innate in indigo bunting infants. A proposal for plausible learning mechanism would be an important first step toward an empirical investigation of whether object representations could be built from perceptual primitives, for such a proposal would certainly make testable empirical predictions. But a proposal for a plausible learning mechanism would not settle the issue.

Two considerations lead me to favor the nativist view that object representations are the product of evolution. First, there is good evidence that the capacity for representation of objects as spatiotemporally continuous and continuing to exist when occluded is innate in chickens (Regolin and Vakkirtugara, 1995). Chicks imprinted on a red ball, who have never in their lives seen any object go behind a barrier (and thus could not have learned about spatiotemporal continuity in the manner by forming generalizations such as 1 and 2 above), act as if they represent the ball as continuing to exist when it goes behind a screen. Of course, that object permanence is innate in baby chickens does not mean it is in human babies. Nonetheless, these studies provide an existence proof that it is possible for the capacity to represent objects as spatiotemporally continuous, even under conditions of occlusion, to be manifest without learning.

Spelke (refs) has best articulated the second consideration that makes me lean toward the nativist position. Spelke has specified principles other than spatiotemporal continuity that constrain young infants’ representations of 3-D objects, and has pointed out that the principles, in concert, determine still other constraints on object interactions. For example, Spelke notes that infants represent objects as bounded and coherent, as well as spatiotemporally continuous. These two principles (spatiotemporal continuity and coherence) entail that one object cannot pass through another. For object A to pass through B, B would have to be noncoherent (like water), or A would have to dematerialize upon hitting B and rematerialize on the other side.

Three separate series of studies indicate that 2-month-old infants look longer at events in which one object has apparently passed through the space occupied by another (Baillargeon et al; Spelke et al., Baillargeon & Hespos, in press), that is, when object motions violate the constraint of solidity. To take just one example, Baillargeon and Hespos (date) showed one group of 2-month-old infants a hollow cylinder and another group an identical cylinder closed on top (Figure 2.8). The cylinder was then placed upright, such that the infant could not see the top. A rod was then picked up and slowly inserted into the cylinder, a possible event in the hollow cylinder case, violating solidity in the closed cylinder case (for the rod would have to pass through the solid top). The infants looked longer at the latter, impossible, event. Recall that 2-months is the earliest age at which infants are adult-like in the Kellman and Spelke amodal completion paradigm (Figure 2.7; Johnson and Aslin, date). It seems that as soon as young infants are able to form object representations, their representations are constrained to reflect
boundedness and spatiotemporal continuity in very complex ways. It is unlikely that piecemeal learning of local statistical regularities could accomplish the coherently interrelated representations observed by 2-months of age.

2.12 Conclusions

I have argued here that the Piagetian/Quinian view of the young infant’s representational capacities being exhausted by a perceptual similarity space, or a set of sensori-motor primitives, is most probably wrong. The argument had three steps. First, I argued that representations of cause and of object cannot be stated in the vocabulary of perception. Second, I reviewed some of the evidence that that young infants represent interactions among objects as causal, and represent objects themselves as spatiotemporally continuous, quantifying over these representations as do adults. Chapters 3, 4 and 7 will present further evidence for these conclusions. Third, I considered whether infants’ performance in the experiments might better be explained in terms of generalizations stated over a perceptual vocabulary. Admittedly, this is still a hotly debated issue, but evidence concerning the inferential role of infants’ representations, together with learnability considerations, lead me to favor the richer interpretation of the currently available data.

Grant, for the moment, that infants’ representations of the world reflect an ontology of individuated, spatiotemporally continuous, middle-sized, middle-distanced, objects that interact with each other according to the laws of contact causality. Grant, contrary to Quine, the
capacity to form and quantify over representations of objects is not a cultural construction, and
does not result from a bootstrapping process that involves learning the quantificational devices of
natural language. Grant, contrary to Piaget, the capacity to represent objects as existing
independently of the child, as spatiotemporally continuous even through occlusion, does not
await the end of the sensori-motor period of development. Granting all this still leaves open
many questions about the origin and development of concepts. What kind of knowledge is this
knowledge of objects and contact causality? Chapter 3 spells out the core knowledge
hypothesis, further differentiating core knowledge from perceptual representations and also
distinguishing core knowledge from intuitive theoretical knowledge. Chapter 4 presents
evidence for additional domains of core knowledge. Chapters 5 and 6 consider the relations
between core knowledge and human language. And finally, chapters 7 through 9 demonstrate
qualitative discontinuities in cognitive development, also sketching the bootstrapping
mechanisms that allow humans to transcend core knowledge.