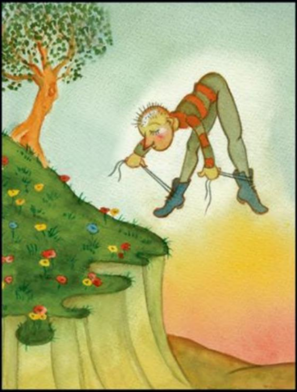
THE ORIGIN OF CONCEPTS



SUSAN CAREY

OXFORD SERIES IN COGNITIVE DEVELOPMENT

The Origin of Concepts

SUSAN CAREY

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
www.oup.com

First issued as an Oxford University Press paperback, 2011

Library of Congress Cataloging-in-Publication Data
Carey, Susan.
The origin of concepts / Susan Carey.
p. cm.—(Oxford series in cognitive development; 3)
Includes bibliographical references and index.
ISBN 978-0-19-983880-6
1. Concepts. I. Title.
BF443.C37 2009
153.2'3—dc22
2008029129

9 8 7 6 5 4 3 2 1

Printed in the United States of America
on acid-free paper

Contents

This page intentionally left blank

The Origin of Concepts

This page intentionally left blank

I

Some Preliminaries

Human beings, alone among animals, come to possess rich conceptual understanding of the world—understanding formulated in terms of such concepts as *evolution, electron, cancer, infinity,* or *galaxy.* This book offers an account of that human capacity for conceptual representation. Different types of processes, operating over three different time courses (individual learning, historical/cultural construction, and evolution), underlie the formation of our conceptual repertoire. Some concepts, such as *object* and *number,* arise in some form over evolutionary time. Other concepts, such as *kayak, fraction,* and *gene,* spring from human cultures, and the construction process must be understood in terms of both human individuals' learning mechanisms and sociocultural processes. Humans create complex artifacts, as well as religious, political, and scientific institutions, that themselves become part of the process by which further representational resources are created.

Cognitive science seeks a precise, explanatory account of the origin of concepts in general, and of some especially important concepts in particular. For example, when it comes to the concept *integer,* what are the relevant innate representational resources bequeathed to us by evolution? In creating such a concept, must we go beyond innately given representations? In what ways, and by what processes, do individuals and groups of individuals construct new representational resources? How is knowledge culturally constructed and maintained? These questions must be answered case by case—there is no reason, at the outset, to expect the answers for concepts of causality or human agency to be the same as the answers for mathematical or biological concepts, or for the concepts that articulate our knowledge of chess or baseball.

This book develops several case studies in some detail: the concepts *object* (chapters 2 and 3), *intentional agent* (chapter 5), *cause* (chapter 6), *integer* (chapters 4 and 7), *rational number* (chapters 9 and 11), and *matter, weight,* and *density* (chapters 10 and 11). In the course of exploring these cases, many others are touched upon. What really matters to me is not the cases (although I do admit finding each one intrinsically fascinating), but the lessons to be drawn from them. My goal here is to demonstrate that the disciplines of cognitive science now have the empirical and theoretical tools to turn age-old philosophical dilemmas into relatively straightforward scientific problems. I shall illustrate the progress science has made in resolving debates about the existence, nature, content, and format of innate knowledge. I consider the thesis that conceptual resources are continuous throughout the life span. I debate the nature of concepts and intuitive theories, and also how to distinguish between conceptual change and belief revision and I show how these controversies bear on the relations between language and thought.

This introductory chapter introduces the problem I seek to solve, states the major theses that are developed in the chapters to come, and provides an overview of the book. As this is a book on the origin of concepts, I begin by sketching what I mean by the term "concept." Concepts are mental representations. Indeed, they make up only part of our stock of mental representations, so discussions of them must distinguish among types of mental representations, saying which ones are concepts. The book's first major thesis is that there are two types of conceptual representations: those embedded in systems of core cognition and those embedded in explicit knowledge systems, such as intuitive theories. The book's second major thesis is that new representational resources emerge in development—representational systems with more expressive power than those they are built from, as well as representational systems that are incommensurable with those they are built from. That is, conceptual development involves theoretically important discontinuities. The book's third major thesis is that the bootstrapping processes that have been described in the literature on the history and philosophy of science underlie the construction of new representational resources in childhood as well.

Concepts and Mental Representations

Concepts are units of thought, the constituents of beliefs and theories, and those concepts that interest me here are roughly the grain of single lexical items. Indeed, the representations of word meanings are paradigm examples of concepts. I take concepts to be mental representations—indeed, just a subset of the entire stock of a person's mental representations. Thus, I use phrases such as "the infant's concept *number*" to mean whatever infant representations (if any) have numerical content, having argued independently that the representation in question (e.g., of number) should be considered a concept, as opposed to a percept or a sensory representation. I assume that representations are states of the nervous system that have content, that refer to concrete or abstract entities (or even fictional entities), properties, and events. I do not attempt a full philosophical analysis of the concept of mental representation itself; I will not try to say how it is that some states of the nervous system have symbolic content. Such a theory would explain how the extension of a given representation is determined, as well as providing a computational account of how that representation fills its particular inferential role, how it functions in thought.[1]

Of course, any theory of the origin of concepts requires some idea of what concepts are and how their content is determined, just as any theory of conceptual content must comport with our best account of how concepts are acquired. Here, I sketch the assumptions that have guided my inquiry from the outset: I assume that there are many components to the processes that determine conceptual content, and that these fall into two broad classes: (1) causal mechanisms that connect a mental representation to the entities in the world in its extension, and (2) computational processes internal to the mind that determine how the representation functions in thought. In broad strokes, then, I assume a dual theory of conceptual content (Block, 1986, 1987).

In order to study the origin of concepts, one must characterize their developmental and evolutionary trajectories, and to do that one must discover what kinds of mental representations and which specific concepts nonhuman animals and human infants and children have. In the pages to come, I work backwards from behavioral evidence for a given representation's extension and inferential role to characterize that representation's nature and content, and whenever possible, to specify something of its format. Chapter 14 draws out some implications for a theory of concepts provided by my picture of their origins.

There are many different types of mental representations, and one challenge faced by cognitive science is to find the principled distinctions among them. Indeed, several historically important views of conceptual development posit shifts in the types of mental representations available to children of different ages, and such claims presuppose distinct kinds of representations. I shall argue that different types of representations may well have theoretically relevant differences in origins, in developmental trajectories, in types of conceptual roles, and in extension-determining mechanisms. Chapter 2 considers one picture of infants' mental lives in which the infant's mind is articulated in terms of representations very unlike those of linguistically and theoretically competent adults. William James famously believed that the infant's world was "one great blooming, buzzing confusion" (James, 1890/1981, p. 496). Quine (1977) suggested that infants begin with a perceptual similarity space that is transformed through learning, especially through learning language, into a system of representations articulated in terms of natural-kind concepts. Similarly, Piaget (1954) proposed that infants begin life with a repertoire of sensorimotor representations, achieving truly symbolic representations only at the end of second year of life.

The views of James, Quine, and Piaget are related in their assumption that a major distinction can be made between sensory representations, on one end of a continuum, and conceptual representations, on the other end. All three theorists posited that sensory/perceptual representations are developmentally primitive, and that uniquely human mental capacities underlie the developmental processes through which sensory/perceptual representations are transformed into conceptual ones. Sensory representations can, of course, be distinguished from perceptual ones, as I shall show in chapter 2, but these writers did not draw a clear distinction between them. For present purposes, the important contrast is between sensory/perceptual representations, on the one hand, and conceptual ones, on the other. All three authors believed that nonhuman animals, just like human babies, do not entertain conceptual representations. Chapter 2 examines the thesis that infants begin with a representational repertoire that is limited to sensory/perceptual representations and that conceptual representations are constructed only later in development, relating this view to the empiricists' theory of the origin and nature of human concepts.

Examining this thesis requires distinguishing the two types of representations (sensory/perceptual vs. conceptual), which is notoriously tricky to do. Modern cognitive science justifies the Empiricists' confidence that representations of the perceptual aspects of the phenomenal world (what things in the world look like, feel like, sound like, smell like, and taste like) result from many levels of processing of the information in physical stimulation impinging on sense organs. That is, as the empiricists knew, there are causal connections between patterns of light, sound, and physical forces and the perception of shape, pitch, depth, and the effort expended in causing something to move. An intuitive first pass at capturing the distinction between perceptual and conceptual representations begins with the observation that representations of such features of immediate experience differ in many respects from clear examples of conceptual representations. The latter represent what things in the world are, categorizing entities into abstract, theory-laden kinds, and positing entities for which we have no sensory evidence.

This first pass at drawing the distinction between sensory/perceptual and conceptual representations presupposes that one can do so on the basis of their content. It assumes one can see, on the face of things, that representations of shapes such as square are sensory or perceptual and that representations of living things or numbers are conceptual. But that can't be quite right, for we certainly have concepts of sensed features of the world. Conflicting with the assumption that shape representations are paradigmatically perceptual, we have words for shapes, develop explicit geometric theories of shapes, and have theories of how shape representations are computed from patterns of retinal input. Conversely, as we shall see, we have perceptual processes that create from sensory input representations of abstractions such as number, agent, and cause.

No doubt there are many dimensions along which clear examples of sensory/perceptual representations differ from clear examples of conceptual ones. Chapter 2 discusses several in the course of exploring the empiricists' proposal (shared by many modern-day cognitive scientists) that the mind begins with a representational vocabulary that is limited to sensory/perceptual primitives. Sensory/perceptual representations are likely to differ from conceptual ones with respect to both important aspects of mental representation—the mechanisms through which their extensions are fixed and the nature of the inferential processes they contribute to. They will also (sometimes) differ in format; I assume many perceptual representations are iconic or analog, whereas at least some conceptual representations are stated over discrete, arbitrary symbols (i.e., are language-like).

One reason for the empiricists' belief that we start out with sensory/perceptual representations and build up concepts from them is that they could imagine how these can be appropriately connected to the world. And they wanted their theory of development to mirror their theory of the determination of content. Although they did not have Darwin to appeal to, they assumed that sense organs guarantee the appropriate causal connections between properties of real-world entities and our perceptual ideas of these entities. I accept this view: the mechanisms of shape perception, for example, are largely innately specified and guarantee that our representations of round things pick out round things. Concepts, for the most part, require different mechanisms to guarantee that a given representation is causally connected to the entities it refers to. Evolution did not provide us with an input analyzer that identifies electrons, justice, or 3,462,179. While there certainly is also a causal story to be told about how the content of the concept *electron* is determined, it will have a very different flavor from the explanation for how the content of *round* is determined. I shall argue, in the course of this book, that inferential role plays a part in the causal processes through which conceptual representations pick out their referents. Additionally, I accept that the principles of content determination for concepts importantly involve social interactions among people and metaphysically necessary features of the entities they refer to. According to philosophers such as Saul Kripke and Hilary Putnam, the extensions of natural kind concepts are fixed not just by the mind but also by some social process of ostensive definition and by the essential nature (a metaphysical matter, not an epistemological one) of the entities so dubbed (Kripke, 1972/1980; Putnam, 1975; see also Burge, 1979). The discovery of the extension of the concept *gold* or of *wolf* is a matter for science, not for philosophy, linguistics, or psychology. As for the psychology of natural-kind concepts, they fall under the assumption of what Susan Gelman calls "psychological essentialism" (Gelman, 2003; Medin & Ortony, 1989). According to the doctrine of psychological essentialism, it is a fact about our mind that we assume (usually correctly, as it turns out, but it needn't be) that individuals of a given natural kind have hidden essences that determine their very existence, their kind, and their surface properties. We assume this even in the face of no guesses as to a kind's essential properties.

In sum, there are many causal processes that are involved in connecting representations in a mind to their referents, and these are likely to differ systematically for sensory/perceptual representations, on the one hand, and conceptual representations, on the other. Whether or not inferential role plays a part in content determination, there is no doubt that perceptual representations and conceptual representations differ in many aspects of their inferential role. Sensory/perceptual representations have different and more impoverished inferential roles from concepts. They represent the here and now, and almost nothing else follows from the fact that something is red, whereas rich inferences are licensed by identifying something as an agent or identifying the substance a given entity is made of. Conceptual representations, but not sensory/perceptual ones, are embedded in conceptual structures such as intuitive theories that support this rich inferential role. Conceptual representations articulate causal and explanatory structures and are integrated with other conceptual representations.

Focusing on conceptual role, the philosopher Jerry Fodor (1983) argued that a systematic distinction between kinds of representations can be drawn in terms of how they are processed in the mind. He distinguished between modular input analyzers and central processes. For the most part, sensory/perceptual processes are modular and conceptual processes are central. Modular processes are data driven, automatic, fast, and encapsulated. (A process is encapsulated if it operates on proprietary input and ignores available information that is relevant to the computation at hand.) Perceptual illusions make the notion clear. Consider the familiar Muller-Lyer illusion (Figure 1.1). You know (because I tell you or because you measured them) that the two horizontal lines are the same length, but you still see the lower one as longer than the upper one. Your perceptual representation is encapsulated from (uninfluenced by) your conceptual representation of the actual relative lengths. Central processes,

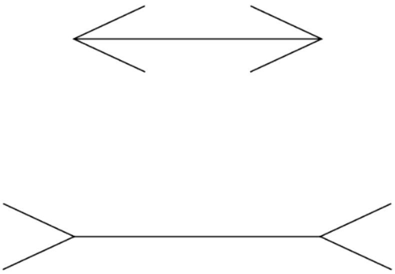


Figure 1.1. The Muller-Lyer illusion illustrates the encapsulation of perceptual processes from explicitly held beliefs. Our *knowledge* of the equality of the lengths of the two horizontal lines does not determine our *perception* of them.

in contrast to modular ones, are informationally promiscuous. That is, there are no limits on what information may turn out, in the end, to bear on any particular hypothesis, and we seek coherence among all of our explicitly held beliefs.

Fodor's processing distinction between perceptual representations (the outputs of modular input devices) and conceptual representations (inferentially interrelated with all other conceptual representations) is important because it highlights the possibility that representations that do not seem sensory/perceptual on their face may nonetheless pattern with them with respect to modularity. His example was syntactic representations. He argued that the parsing processes that output abstract linguistic representations that are formulated in terms of concepts such as *noun phrase* and *verb phrase* are modular, and thus such representations are more like sensory/perceptual ones than conceptual ones.

Thesis 1: Core Cognition

This book's first major thesis is that there is a third type of conceptual structure, dubbed "core knowledge" by Elizabeth Spelke, that differs systematically from both sensory/perceptual representational systems and theoretical conceptual knowledge.[2] Here I use the locution "core cognition" rather than "core knowledge" because the representations in core cognition need not be (and often are not) veridical and therefore need not be knowledge. I shall argue that nonhuman animals, as well as human beings, have systems of core cognition, and that core cognition is the developmental foundation of human conceptual understanding. Like sensory and perceptual features of the world, the entities in core domains of knowledge are identified by modular innate perceptual-input devices. Therefore, the extension of the symbols that articulate core cognition is fixed, in part, by evolutionarily underwritten causal relations between entities in the world and representations in the mind. However, representations in core cognition differ from sensory and perceptual representations in having a rich, conceptual, inferential role to play in thought, even infants' thought. Representations that are the outputs of distinct core cognition systems are inferentially integrated and are in this sense central.

Still, core cognition representations differ from the fully explicit conceptual representations that articulate intuitive theories. The conceptual role of the concepts that articulate core cognition is vastly less rich than that of the concepts embedded in intuitive theories. They also differ in the mechanisms that connect them to the entities they represent. Other factors, in addition to causal connections mediated by innate perceptual-input analyzers, play a role in determining the referents of fully conceptual representations. These include the social processes described by philosophers such as Kripke, Putnam, and Burge. I shall argue that the inferential role has a crucial part to play in the processes through which the referents of conceptual representations are determined. Finally, knowledge acquisition in core domains is supported by innate domain-specific learning devices, whereas the learning of intuitive theories is not.

At Stake, A Picture of Conceptual Development

The thesis that core cognition exists is a nativist claim. Some authors have doubted that it would be possible for there to be innate representations, or that the notion is even coherent (e.g., Elman et al., 1996; Thelen, Schoner, Scheir, & Smith, 2001). Authors have different notions in mind when they use the term "innate," so it is important to be clear. As I use the term, innateness is interdefined with learning: innate representations are those that are not the output of learning processes. This may seem unsatisfying, for it simply pushes the problem of understanding what innateness is onto the problem of understanding what learning is. Nonetheless, I believe that this contrast is both coherent and that it matters in the debate about which representations are innate. This book is an extended exploration of learning mechanisms, but for present purposes, we can characterize learning processes as those that build representations of the world on the basis of computations on input that is itself representational. Clear examples are mechanisms that chose among alternative hypotheses on the basis of rational processes that weigh evidence (as in Bayesian learning processes), operant conditioning, connectionist supervised and unsupervised learning algorithms, habituation, and so on.

Notice that claiming that representations of red or round are innate does not require that the child have some mental representation of red or round in the absence of experience with red or round things. The capacity for forming color or shape representations could be innate (i.e., not constructed through learning), even though no representations of colors or shapes are ever activated until entities are seen. Notice also that "innate" does not mean "present at birth." Many representational capacities arise from maturational processes. An example is stereoscopic representations of depth, which emerge in humans quite suddenly around 6 months of age. Even though stereoscopic depth perception is not present at birth, I would want to say it is innate, for the child does not have to learn to compute depth from the discrepancies between the two images on the two retinas. Another reason some innate (unlearned) representations may not be evident at birth is that the child has not yet encountered the input to the processes that construct them or may not be yet able to represent their inputs. Chapters 2 through 5 provide examples of this sort. For a sophisticated discussion of the concept of innateness as it applies to mental representations, see Spelke and Newport (1998).

I take it that any theory of conceptual development must specify the stock of innate representations, as well as the mechanisms (both maturational and learning mechanisms) that underlie change. Framing the problem this way leaves all the room in the world for cantankerous disagreement. Are the initial representational primitives limited to sense data, as the British empiricists such as Locke would have had it? Or to sensori-motor schemes, as Piaget would have had it? Or to a perceptual similarity space, as Quine would have had it? Or is it theories all the way down, as Allison Gopnik and Andrew Meltzoff would have it?[3] Insofar as learning underlies conceptual development, are the learning mechanisms domain-general or specifically tailored to particular conceptual content? If domain-specific, what are the domains and what are the mechanisms of learning within each? If domain-general, are they powerful pattern abstractors, such as those well modeled in connectionist systems, determinate computational algorithms formulated over symbolic machinery, or bootstrapping mechanisms that draw on nondeterminate processes such as analogical mapping, inference to best explanation, and inductive guesses?

The proposals for the initial stock of representational primitives and for the types of learning mechanisms that underlie cognitive development are logically independent, although there are inductive relations between them. Typically, researchers who believe that the initial state of the baby consists of sensory representations alone also believe that domain-general learning mechanisms of various sorts subserve conceptual development. This position (innate sensory representations/domain-general learning mechanisms) is discussed in chapter 2. A second view (innate conceptual representations/domain-general learning mechanisms), held most forcefully by Gopnik and Meltzoff (1997), is that the initial stock of representational resources includes conceptual representations embedded in intuitive theories and that development is driven by domain-general causal learning mechanisms of the sort that underlie theory development in science. I endorse a version of this position, but not with respect to the characterization of the initial state. Chapter 3 contrasts this position with a third position (innate conceptual representations/domain-specific learning mechanisms), which holds that the initial stock of representational resources includes conceptual representations embedded in core domains and that development is driven, at least in part, by domain-specific learning mechanisms.

It is important to note that the existence of innate conceptual representations and domain-specific learning mechanisms does not preclude the existence of sensory and perceptual representations and domain-general learning mechanisms, or even that there are learning mechanisms capable of taking perceptual representations as input and outputting conceptual ones. Building rich and accurate representations of the physical, social, and biological worlds is so important—to humans especially, but to all animals really—that many distinct representational and learning mechanisms are likely to have been selected for. As Gallistel (2000) commented, speaking of the proposal that there only one learning mechanism, "From a biological perspective, this assumption is equivalent to assuming that there is a general purpose sensory organ that solves the problem of sensing."

A variety of considerations lead individual scientists to favor one or another picture of the initial state and of the processes that yield adult conceptual capacities. Some reduce to matters of scientific strategy, such as the dictum that it is scientifically preferable to assume the least possible in the way of innate representations, in order to explore how much of the adult state could be accounted for under any particular minimal assumptions (e.g., sensory primitives, associationist learning mechanisms). But scientific strategy must not be promoted to principles of theory choice with truth-relevant status, as when it is claimed that such a picture is more parsimonious than alternative pictures and thus should be considered right until proven wrong.

Other scientific strategies lead to a different picture. For example, some scientists bet that human learning processes will be continuous with those of other animals—an assumption that generates an expectation of highly domain-specific and structured learning mechanisms. There are myriad examples of specialized learning mechanisms in other animals—ranging from passerine song learning, to vervet learning to identify predators, to spatial learning in rats, to migratory birds learning the azimuth of the night sky (Gallistel, Brown, Carey, Gelman, & Keil, 1991, review these particular cases). In each of these cases, innately specified representations and domain-specific learning processes are essential parts of the mechanisms that achieve the adult state.

Two Examples of Animal Core Cognition

I now turn to two examples from the ethological literature that show that it is not problematic to attribute innate representations and domain-specific learning mechanisms to nonhuman animals. These examples illustrate what is meant by core cognition, and they raise the question: What core systems has evolution endowed humans with?

Indigo Buntings Learn to Identify North from the Night Sky

Consider the mechanism through which indigo buntings (a species of migratory songbird) learn to identify north from the night sky. In today's night sky, Polaris is the North Star, the star in the region of the sky that reliably indicates north. Because the stars change position over time, Polaris will not always indicate true north, nor will true north in the future be predictable from the constellations of today, for the constellations themselves migrate and change over evolutionary time. Indeed, 100,000 years ago there was no Big Dipper, no Orion. Thus, evolution probably could not have built a map of the sky in the brain of the indigo buntings on which Polaris is identified as north, yet indigo buntings have such a map. Presented with a stationary simulacrum of the night sky in a planetarium in autumn, they take off as if to fly south as specified by the North Star, no matter how the planetarium's night sky has been oriented with respect to true north in the actual world. Indigo buntings must have learned something equivalent to where the North Star is; how could they have done so?

Steven Emlen (1975) uncovered the mechanism by which nestling indigo buntings achieve this feat. Because the earth rotates on an axis that goes through true north, the North Star, positioned as it is at present above the North Pole, marks the center of rotation for the night sky. Emlen showed that nestling indigo buntings observe the rotating sky and register the center of the rotation as a privileged direction. He showed this by raising indigo buntings in a planetarium in which he could make arbitrary skies rotate around arbitrary centers, and then observed which direction the birds took off in the autumn, when their hormones told them it was time to fly south. There is a critical period for this learning device; if the buntings do not learn the North Star while they are nestlings, they never do.

This is a paradigm species-specific, domain-specific learning device. Clearly, not every animal will spontaneously note the center of rotation of the night sky, nor will every animal infer true north from the observations if they happen to make them; thus, this device is species-specific. With respect to domain specificity, this device is of no use in learning what food is safe to eat, what the indigo bunting song is, or anything else other than where north is. This device exemplifies other properties of core domains, as well. Innate perceptual mechanisms allow the bird to identify the sky and to represent its rotation. And this device is a learning mechanism; innately specified computations concerning the representation of celestial rotation ensure that the buntings end up representing something important that they must, by logical necessity, learn: how to tell north from the night sky.

That some state of the indigo bunting's nervous system represents the night sky is shown not only by the fact that it is activated by the night sky but also how it plays a role in the computations that determine the azimuth and direction of flight during migration. This is an example of what I mean by taking behavioral evidence regarding the extension and inferential role of some representation as evidence regarding its content. Finally, these representations also exemplify the senses in which core cognition is conceptual, but not fully so. The inferential role of the representations of a view of the night sky go beyond the specification of what it looks like (points of light distributed against a black background); nothing in the description of what the sky looks like has the content north. Still, the inferential role of bunting representations of the azimuth is limited (as far as we know) to guiding navigation. In spite of a lack of studies on the matter, it's a safe bet that buntings do not create theories of the earth and the heavens in terms they can use to explain why the center of rotation of the night sky provides information about compass direction on the earth. Thus, it would be a mistake to suppose that the bunting's representations of the night sky are anything like ours.

Conspecific Recognition

Ever since the pioneering work of Konrad Lorenz (1937), we have known that many birds have an innate learning mechanism that enables them to identify conspecifics, and that these representations are input to a learning process that enables them to identify one important conspecific: their mother. The learning device Lorenz discovered and called "imprinting" causes chicks to attend to and seek proximity to the first large thing that moves in certain ways. Countless beginning psychology and ethology students have seen the footage of Lorenz walking through the countryside with a line of goslings imprinted on him, following behind. In this case, the perceptual analyzer that initially identifies mother (or conspecific) does so on the basis of movement of a certain sort.

Goslings and chicks will imprint on a large red shiny ball if that is the only object they see moving during a critical period after birth—or so we thought until the work of Mark Johnson and his colleagues (M. Johnson, Bolhuis, & Horn, 1985). These researchers showed that mother/conspecific identification is such an important problem that evolution has provided at least two redundant learning devices to support it. A newborn chick presented with a moving ball or person and also a stationary stuffed hen will huddle near the stuffed hen.

Thus, Johnson and his colleagues demonstrated that chicks must have an innate perceptual analyzer that specifies what a conspecific looks like. Indeed, these researchers specified the nature of these innate representations. The innate representation of conspecific appearance is very sketchy—there must be an overall bird shape, with a head with eyes and a beak on a neck on a body; if these elements are not in the right configuration, the chick does not recognize the object as a conspecific. However, a chick will huddle close to a stuffed duck or eagle or owl. Johnson and his colleagues have much to say about this fascinating story, bearing on the relations between the two different learning mechanisms, their critical periods, and their neural substrates. My point here is simply that there is an innate representation of hen that specifies roughly what hens look like and contains the inferential role "stay close to that." This representation is part of a domain-specific learning device; the representation is sketchy so it can be filled in with the details of the chick's own mother. In this way it is similar to the innate representations human babies have of conspecifics (two eyes above a mouth within an oval, in a certain configuration—representations that play a role in human babies' identifying others like them and learning to recognize individual caregivers; Morton, Johnson, & Maurer, 1990).

This book argues for many controversial conclusions, but the claim that some representations are innate is not meant to be one of them. The controversy comes when we begin to consider the content and nature of innate representational systems. There certainly are innate sensory representations. Also, innate input analyzers, constrained to accept only some classes of stimuli (bird-shaped entities, face-shaped entities, the night sky), create representations that clearly go beyond sensory content and underlie learning about entities in the world that are crucial to individual survival.

Thesis 2: Discontinuities in the Course of Conceptual Development

This book's second major thesis is that human beings, alone among animals, have the capacity to create representational systems that transcend sensory representations and core cognition. That is, human beings create new representational resources that are qualitatively different from the representations they are built from. Many cognitive scientists, such as Jerry Fodor (1975, 1980) and John Macnamara (1986), deny the very possibility of true cognitive development in this sense, endorsing instead a strong version of the continuity thesis. The continuity thesis is that all the representational structures and inferential capacities that underlie adult belief systems either are present throughout development or arise through processes such as maturation.

Fodor's (1980) argument for the continuity thesis was a one-liner: one cannot learn what one cannot represent. In the famous debate between Piaget, on the one hand, and Fodor and Chomsky, on the other (Piatelli-Palmarini, 1980), Fodor argued that all known learning mechanisms (e.g., parameter setting, correlation detection, prototype formation) are variants of hypothesis testing, and one cannot confirm a hypothesis unless one can already represent it. That is, learning involves choosing between already formulated hypotheses, setting a previously specified parameter, noting a correlation between already represented states of affairs, and abstracting what is common among a set of represented exemplars. These mechanisms, by their very nature, cannot yield a capacity to represent anything that was previously unrepresentable.

Fodor (1975) is famous for an application of this general argument, leading to the conclusion that all lexical concepts are innate. This conclusion requires extra assumptions about lexical concepts—namely that they are atomic, or can't be decomposed into primitives. I eventually address this version of Fodor's continuity thesis in chapter 13, but for now my concern is the more general argument. I do not contest the truism that all learning involves building new representations from antecedently available ones, nor that that all learning in the end can be analyzed as forms of hypothesis fixation. My concern is where the hypotheses come from. As I demonstrate in chapters 8 and 11, learning processes exist that can create representations more powerful than their input. These then allow hypotheses to be formulated over concepts that cannot be expressed in the vocabulary and syntax available at the outset of the learning episode.

Fodor's argument has a crucial weakness. It is falsified by a single counterexample that shows that novel representational capacities, with more expressive power than their input, must have been formed through some learning process. Piaget (1980) invoked an excellent set of counterexamples in his own reply to Fodor. Drawing from the history of mathematics, Piaget pointed out that concepts of rational, real, and complex numbers, and the mathematical notations required to express them, cannot plausibly be attributed to infants, children, or even adults lacking sufficient education in mathematics, and they do not plausibly develop through maturation. Rather, such concepts arise as one learns mathematics. Fodor might reply that infants must innately have the representational resources to express these concepts, but there is an ambiguity in that reply. Of course, they innately have the capacity to construct those representational resources—that is not in dispute. What's actual is possible, and many mathematically literate people represent rational numbers. The question is whether infants could express concepts such as *one-third* or *pi*. I take these questions up in chapters 8 and 9.

Chapter 8 shows that the issue arises even in the case of the positive integers. The integer list is a cultural construction with more representational power than any of the core representational systems on which it is built, thereby providing a genuine counterexample to the argument that conceptual discontinuities are in principle impossible. Chapter 9 sketches the later (in history and in ontogenesis) construction of the concept *rational number,* which equally well transcends the representations available at the outset of the construction process (namely, representations of integers created by children during their preschool years). In these cases, discontinuity is cashed out in terms of vast increases in expressive power.

The study of cognitive development provides many further counterexamples to the continuity thesis. In the course of acquiring intuitive theories of the world, children create new concepts that are incommensurable with those from which they are built. In such cases, discontinuity is cashed out in terms of the creation of new concepts not translatable into the concepts available at the outset of the episode of conceptual change. Chapters 10 and 11 discuss case studies of conceptual change in the history of science and also in childhood.

If such existence proofs show that Fodor's argument is wrong, however, they do not show what is wrong with it. In particular, they do not show us how the infant, child, mathematician, or scientist can use his or her current representational resources to learn new ones. That is the real challenge to cognitive science raised by Fodor's argument, and it is still unmet.

Those who deny the continuity thesis actually face two distinct challenges, one descriptive and one explanatory. First, descriptively, we must provide a satisfactory characterization of what it means for a representational system to be qualitatively different from, to transcend, those that preceded it. We must also provide evidence for two successive points in cognitive development, the later of which contains a representational capacity that transcends what was previously available. Second, with respect to explanation, we must then specify a learning mechanism that accounts for how new representational capacity could come into being.

Thesis 3: Quinian Bootstrapping

This book's third important thesis is that the explanatory challenge is met, in part, by bootstrapping processes such as those described in the literature on history and philosophy of science. I call the kind of bootstrapping process we need to understand conceptual discontinuity "Quinian bootstrapping" because Quine (1960, 1969) developed particularly colorful metaphors in trying to explicate the idea (e.g., building a chimney, pressing against the sides to support oneself as one scrambles up it, building a ladder, and then kicking the ladder out from under, and Neurath's boat, in which one builds a structure to support oneself while already at sea). To "bootstrap" means, literally, to pull oneself up by one's own bootstraps—something that is clearly impossible. The metaphor captures what is hard about the process of creating new representational resources that are not entirely grounded in antecedent representations. Chapters 8 and 11 offer an account of the bootstrapping metaphors of historians and philosophers of science in terms of the resources from current cognitive science.

The Quinian bootstrapping that underlies discontinuous conceptual development must be distinguished from a different kind of learning, also called bootstrapping, that is debated in the contemporary literature on language acquisition. In the language-acquisition literature, bootstrapping processes are invoked to explain how children solve a mapping problem. Suppose children know, thanks to innately supported universal grammar, that there will be nouns, noun phrases, transitive verbs, verb phrases, adjectives, prepositions, quantifiers, and so on, in natural language. Suppose, moreover, that they create innately supported representations for individuals and kinds of individuals for actions, stuffs, intentional causation, and so on. Children still face the formidable problem of identifying how their own particular language expresses these and other universal features of language and thought. Semantic bootstrapping—the use of semantic information to infer syntactic categories—gives the child a beginning wedge into the problem of discovering a particular language's syntactic devices. For example, the language learning mechanism might include the heuristic that representations of kinds of physical objects ought to be mapped onto count nouns. Then, if the child can figure out that a word is being used to refer to a kind of object, the child may assign it to the lexical category count noun and use this assignment to figure out how count nouns are marked in their language (see, for example, Pinker, 1984). Syntactic bootstrapping exploits innately given initial mappings of this sort to solve the converse problem—that of discovering that particular words express particular concepts. For example, if the concept *give* includes a giver, a receiver, and a gift, then the child may exploit syntactic evidence that a given verb has three arguments to guess that it might express this concept (see, for example, Gleitman, Cassidy, Napa, Papafragou, & Trueswell, 2005). Both syntactic and semantic bootstrapping require antecedent conceptual and linguistic representations, and they support solving a very difficult mapping problem. However, these bootstrapping processes do not deny continuity—no totally new representational resources are being created.

Some earlier students of language acquisition also appealed to Quinian bootstrapping as a process underlying language acquisition. Isaac Schlessinger (1982), for example, hypothesized that initially linguistic elements receive only semantic interpretations, such that creating syntactic categories required constructing representations previously unrepresentable.

Rather than just solving a mapping between previously available types of representations, the process envisioned by Schlessinger creates a new type of representations. That is, he suggested that syntactic categories are created by a bootstrapping process of the sort envisioned in the history of science literature (see also Braine, 1963). It is beyond the scope of this book to consider whether Quinian bootstrapping is needed in the course of syntactic development—at issue in that debate is how rich the innate language acquisition device is. But I will show that Quinian bootstrapping definitely has a role to play in explaining the origin of concepts.

1.7. Intuitive Theories—Explicit Conceptual Representations

As I mentioned at the beginning of this chapter, I distinguish two types of conceptual representations: those that articulate core cognition and those that articulate later developing linguistically encoded knowledge structures like intuitive theories. Intuitive theories differ from core cognition in each of its distinctive features: they are not innate, the entities in their domain are not identified by innate input analyzers, their format is most probably not iconic, and they are not continuous throughout development. Quinian bootstrapping mechanisms underlie the human capacity to create theoretical knowledge that transcends core cognition.

Although there are many knowledge structures worthy of study (scripts, schemas, prototypes, the integer list representation of number, the alphabet), many students of cognitive development assume that one kind of knowledge structure—intuitive theories—plays a particularly important role in cognitive architecture. I endorse this assumption, and here I focus on a class of intuitive theories—those that Henry Wellman and Susan Gelman call "framework theories" (Wellman & Gelman, 1992; see also Carey, 1985; Gopnik & Meltzoff, 1997; Keil, 1989). These are the theories that ground the deepest ontological commitments and the most general explanatory principles in terms of which we understand our world. One task (but by no means the only task) in the study of cognitive development is to account for the acquisition of framework theories.

It is worth stepping back and considering what is being presupposed by the choice of the term intuitive theory rather than the more neutral cognitive structure. Intuitive theories, like scientific theories, play several unique roles in mental life. These include: (1) determining a concept's most important features (the properties seen as essential to membership in a concept's extension); (2) representing explicitly held causal and explanatory knowledge; and (3) supporting explanation-based inference. Furthermore, the mechanisms underlying theory development, including Quinian bootstrapping, differ from those that underlie the acquisition of different types of conceptual structures. It is an empirical question whether children have intuitive theories, and whether knowledge acquisition in childhood involves the process of theory change. Those who talk of "intuitive theories" and "framework theories" are explicitly committing themselves to an affirmative answer to those empirical questions. This commitment does not deny that there are important differences between children as theorizers and adult scientists (hence the qualifier, intuitive). Children are not metaconceptually aware theory builders (D. Kuhn et al., 1988). In spite of these differences, the research enterprise in which this work is placed presupposes that there are important substantive similarities between scientific theories and intuitive theories. Of course, the merit of this presupposition depends on the fruitfulness of the research it generates. Chapters 10 and 11 present case studies framed within this research tradition.

As mentioned above, intuitive theories differ from core cognition in many ways. One of the goals of this book is to account for the origin and development of theory-embedded conceptual knowledge, given a beginning state of perceptual representations and core cognition. Many researchers have blurred the distinction between core cognition and intuitive theories. For example, Alan Leslie is one of the most articulate advocates of the core cognition position, and he—like me—characterizes core cognition as modular, encapsulated, supported by innate perceptual analyzers, and unchanging during development. Confusingly, in discussing the infants' beginning state, he characterizes core cognition in much the way I have, yet he dubs his modules Theory of Mind Module and Theory of Bodies Module (Leslie, 1994). That is, he characterizes core cognition, but he calls systems of core cognition "theories." This is not merely a squabble about terminology, for if I am right that core cognition exists and differs from explicit conceptual knowledge in the ways specified above, then failing to adopt contrasting terminology conflates fundamentally different kinds of mental representations.

Gopnik and Meltzoff (1997) go further than merely calling infant conceptual knowledge "theories"; they deny the distinction between core cognition and theories. They explicitly characterize infants' early developing knowledge of bodies and agents as theoretical knowledge, claiming that the mechanisms that underlie acquisition of knowledge in these domains are the same as those that support theory development by adult scientists. This leads to the seemingly absurd (and false, I shall argue) conclusion that the processes by which infants achieve object permanence are the same as those through which Darwin formulated the theory of natural selection.

The confusion between core cognition and intuitive theories is intelligible, for two of the paradigm cases of each have overlapping content. Core cognition of objects overlaps with knowledge of intuitive physics, and core cognition of intentional agents overlaps with an intuitive theory of mind. This overlap in content arises because the output of the core cognition systems is part of the input to theory building. The case studies in this book constitute an extended argument for the position that theory building can, and does, transcend core cognition.

Overview of the Book

The chapters to come distinguish the representations that constitute core cognition from two other types of representations: sensory/perceptual representations and conceptual representations embedded in explicit intuitive theories. Chapter 2 examines the hypothesis that the developmental primitives are sensori-motor or perceptual. Chapters 3, 4, and 5 characterize core cognition and provide evidence for several domains of human core cognition that exemplify its distinctive features. Chapter 6 considers the question of whether all innate representations with conceptual content are embedded within core cognition systems, or alternatively, whether there are also innate central representations. Chapter 7 touches on some relations between core cognition and language. Chapters 8 through 11 take on both parts of Fodor's challenge to cognitive science. They provide descriptions of discontinuous conceptual development in which concepts come into existence that were not expressible given earlier conceptual resources, and they characterize the bootstrapping mechanisms that underlie the change. Finally, the two concluding chapters summarize my account of the origins of concepts (chapter 12) and draw out the implications from these case studies for a theory of human concepts (chapter 13).

NOTES

1. An excellent overview of competing accounts of concepts in the philosophical and psychological literatures is provided in the collection of classic papers assembled by Margolis and Laurence (1999). See also their own superb and comprehensive tutorial essay in the same volume.

2. See Carey & Spelke, 1994; R. Gelman, 1991; Leslie, 1994, and Spelke, Breilinger, Macomber, & Jacobsen, 1992, for related characterizations of core cognition.

3. Different pictures of the initial representational repertoire have articulated theories of conceptual development in the Western philosophical/scientific traditions at least back to the time of Greek thinkers. Here I am referring to works by Gopnik & Meltzoff, 1997; Locke, 1690/1975; Piaget, 1954; Quine, 1960, 1977.

This page intentionally left blank

2

The Initial Representational Repertoire: The Empiricist Picture

In this chapter I begin to build the argument for core cognition by marshaling evidence against the empiricist hypothesis that the initial state is limited to sensory representations. Many modern thinkers, as disparate as Piaget, Quine, eliminitivist connectionists, and systems dynamic theorists explicitly or implicitly share this empiricist assumption (e.g., Elman et al., 1996; Piaget, 1954; Quine, 1960, 1977; Thelen et al., 2001). This hypothesis has a certain plausibility. All evidence we have of the particular world we live in comes through our senses. Doesn't this mean that all of our knowledge must be able to be formulated in terms of sensory primitives?

Here I counter this seductive argument, while summoning evidence that some early developing representations are conceptual. The chapter has some secondary goals as well. I discuss the methods I draw upon throughout the first half of this book. Also, different methods yield apparently conflicting data, and I give some sense of how conflicts might be resolved. Finally, I lay out the arguments that convince me that some conceptual representations are innate.

The Empiricist Picture

According to British empiricists such as John Locke, all human concepts are grounded in a set of primitive representations—in Locke's terms, "ideas." The primitive ideas are the output of sense organs—they are sensory representations. They are primitive in two different senses. First, these ideas are definitional primitives. All concepts are either primitive or complex, and all complex concepts are defined in terms of primitive ones that themselves are understood without any definition (Locke, 1690/1975; see Margolis & Laurence, 1999, for a superb tutorial on the empiricist theory of concepts). Second, these ideas are developmental primitives. 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.

The 18th-century British empiricists' picture of conceptual development finds articulate and ardent defenders to this day. This staying power has two explanations. First, 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. They offered a comprehensive theory that provided an integrated account of all of these phenomena and more. Few contemporary theories of concepts have anything like the scope of the empiricists'. Second, the theory contains important grains of truth.

That the primitive ideas are sensory was important to the empiricist program. The empiricist explanation of reference depended on the view that sensory representations refer to certain aspects of the world by virtue of the operation of the sense organs. That is, they took as unproblematic the view that our concepts red and round are based on the cases in which red and round things cause us to have sensory representations of red and round. As long as the referential potential of primitive concepts is guaranteed by how sensation works, and as long as all concepts may be defined out of primitive concepts, then the referential potential of all concepts is explained. That is, the extensions of complex concepts are determined by their definitional structure and the extensions of the primitives from which they are built.

This is not the place for a full exposition and critique of the empiricist view of concepts. Every part of the view has come under fire (see chapter 13). Most obviously, the project of defining all concepts in terms of sensory primitives is unworkable. Most representations underlying human natural language are not perceptual representations. Human beings represent nonobservable entities (beliefs, protons), nonobservable properties of observable entities (functions, essences), abstract entities (numbers, logical operators), and fictional entities (Gods, ghosts, Hamlet). Concepts for such entities are not themselves the output of sense organs. Of course, the empiricists held that these concepts are nonetheless definable in terms of perceptual primitives. No adequate definition has ever been provided for most concepts (Fodor, Garrett, Walker, & Parkes, 1980; Laurence & Margolis, 1999), and certainly no definitions for many concepts in terms of perceptual primitives can even be attempted. Try doing so for the concepts *justice* or *sin*, let alone the concept *God*.

This line of argument defeats the position that all human representations are either perceptual or defined in terms of perceptual representations. However, it does not defeat the argument that the innate primitives are perceptual, as long as one provides a learning mechanism that could account for the creation during development of nonperceptual representations, given a beginning stage containing only perceptual representations. I believe that the bootstrapping processes that are characterized in chapters 8 through 11 underlie discontinuities in conceptual development, and thus I do not accept a general learnability argument for the impossibility that the initial state may consist only of perceptual, sensori-motor representations. Rather, if we wish to characterize the innate representational primitives, we have no alternative but to do the hard empirical work of finding out what representations young infants have.

The empiricists certainly got two important points right. There are innate sensory representations and their content is ensured by how our sense organs work. The empiricists did not know that they could thank natural selection for making this is so, but Darwin gave us a way of understanding how the right causal connections between properties of the world and states of the nervous system can be established and maintained. It is because sense organs were selected to work as they do that humans can see color and movement, can taste salt and sweet, can hear tones and feel heat. The operation of evolutionarily constructed input analyzers guarantees that the relevant representations refer to aspects of the environment that are important to survival.

A Historical Aside: The Rationalist/Empiricist Debate About Perception

Sensory representations may be roughly characterized as those representations that are the output of the sense organs. They are what psychologists call proximal representations—those representations that maintain the point of view of the pattern of stimulation on sense organs. For example, a retinal projection is a proximal representation, as is a representation of a pin prick on the back of my hand. As has been well known for centuries, the information in the proximal stimulus greatly underspecifies the distal world that was the source of the stimulation in the first place. Distal representations capture aspects of the external world—they exhibit constancies. Perceived size compensates for the fact that proximal representations of the sizes of objects are determined by distance; the retinal projection of a quarter is much larger at arms' length than 5 feet away, yet our perceptual representation of the actual size of the quarter is accurate over these distances. Similarly, perceived shape is three-dimensional, in spite of the fact that retinal projections are two-dimensional.

Because the empiricists were committed to the view that developmental primitives were sensory, they faced a formidable challenge in accounting for our capacity to perceive the true shapes of objects, the true depth relations among them, and so on. One arena of the historical debates between empiricists such as Berkeley and Hume and rationalists such as Descartes was their solutions to this challenge (e.g., Berkeley, 1732/1919; Descartes, 1637/1971[1]). Consider the representation of depth as an example. Empiricists such as Berkeley and Hume attempted to show how representations of depth could be learned by associative mechanisms operating over sensory primitives. They considered sensory primitives such as proximal representations of size and shape, interposition cues, convergence of the eyes (felt effort in the muscles being greater as a function of how converged the eyes are), and accommodation of the lens (also conceptualized as felt effort). They then assumed that these cues are associated, through learning, with cues to depth from other sensory modalities—for example, proprioceptive cues to the difference between reaching and contacting a nearby versus a more distant object, or between walking to a nearby object and walking to a more distant object. According to the empiricists, the veridical representation of depth is built up from and constituted by this associative structure.

Nativists such as Descartes presented a very different picture. Descartes did not deny that the information from which depth representations were computed must be the output of sense organs. Nativists do not believe in magic. Rather, he believed that there are innate inferential mechanisms (today we would say computational mechanisms) that instantiate constraints derived from geometry that take this input and transform it into veridical representations of depth. His example was the geometrical inference from convergence to depth, an argument ridiculed by Berkeley (1732/1919) thus:

> But those lines and angles, by means whereof mathematicians pretend to explain the perception of distance, are themselves not at all perceived, nor they, in truth, ever thought of by those unskillful in optics. I appeal to any one's experience, whether, upon sight of an object, he compute its distance by the bigness of the angle made by the meeting of the two optic axes? Or whether he ever think of the greater or lesser divergence of the rays, which arrive from any point to his pupil? Nay, whether it be not perfectly impossible him to perceive by sense the various angles wherewith the rays, according to their greater or lesser divergence, do fall on his eye. Every one is himself the best judge of what he perceives, and what not. In vain shall all the mathematicians in the world tell me, that I perceive certain lines and angles which introduce into my mind the various ideas of distance; so long as I myself am conscious of no such thing. (pp. 15–16)

Of course, we now know that Descartes was entirely right—exactly those computations are instantiated in the mind and they do contribute to the perception of depth. Modern models of perception do not require that the representations that enter into modular perceptual computations be consciously accessible. Another example in the tradition of Descartes would be the computation of depth from information derived from the degree of mismatch between the images of an object derived from each eye (stereopsis). Berkeley could similarly ridicule this idea by saying that he is not aware that the two eyes yield different proximal images of an attended object, and that even if he were aware of this fact, he wouldn't know how to calculate depth from it.

As far as I can tell, there is hardly any classical debate from the history of philosophy of mind that has been more conclusively settled. Every textbook on perception details the computations carried out over sensory representations that yield veridical representations of depth, and all agree that such computations are subconscious, operate on proprietary information, and are encapsulated (i.e., are modular, in Fodor's sense, at least to some degree). Showing Berkeley wrong, the evidence that at least some of these computations are innate is overwhelming. Eleanor Gibson's famous work on the visual cliff provided some of the first evidence in support of innate mechanisms for depth perception: newborn animals who have had no opportunity to form associative structures over different sensory cues to depth avoid the deep end of a visual cliff (Gibson & Walk, 1960). With respect to human infants, Alan Slater's demonstrations of size constancy in newborns require that the infant represent depth (Slater, Mattock, & Brown, 1990).

I draw two morals from this story. First, the question of whether development begins with a stock of merely sensory primitives, or whether evolution endowed us with computational devices that yield veridical representations of the distal world, is settled in favor of the existence of Descartes-like innate perceptual input analyzers. Second, there is no in-principle argument against the hypothesis that evolution endowed animals with input analyzers that yield representations that are further along the continuum between sensory representations and conceptual ones than are depth representations. Representations further along this continuum will be couched in the vocabulary of abstractions rather than that of appearances and spatio-temporal relations. They will be central, interacting with the output of other input analyzers, will be accessible, and will have relatively rich inferential roles.

The Initial State: Perceptual/Sensori-motor Primitives

Important 20th-century psychologists and philosophers, as different as Jean Piaget (1954) and W. v. O. Quine (1960), also held that the initial repertoire of mental representations is limited to a set of sensory or perceptual developmental primitives. Piaget's position was that infants begin life solely with representations that subserve innate sensori-motor reflexes. All mental life, according to Piaget, is constructed from this initial representational repertoire. For Piaget, the important properties of sensory representations that distinguished them from conceptual ones included their being the output of sense organs (and restricted to single-sense modalities) and their content being limited to currently experienced sensations.

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 analytic definition made sense; he denied the analytic/synthetic distinction). Nonetheless, he held that the infant's initial representational resources were limited to an innate perceptual vocabulary, which he called a "prelinguistic quality space" and which he conceptualized as an innate perceptual similarity space. In a series of influential writings Quine developed three interrelated theses about conceptual development (Quine, 1960, 1969, 1977):

1. Infants' representations are radically different from those of their elders, and are formulated with respect to 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.

This book is an extended meditation on these three theses. In these early chapters I argue, contrary to Quine, that many infant representations are conceptual and that many of our commonsense ontological commitments are innate. However, I agree with Quine that some, indeed most, of our commonsense ontological commitments are a cultural construction, and in chapters 8 through 12 I will spell out how the bootstrapping processes he envisioned work.

Whether there are innate conceptual representations is an empirical question. Because both Piaget and Quine focused their discussions on representations of objects (as did the British empiricists), I begin with a case study of infant object representations. Both Piaget and Quine agreed that young infants cannot achieve representations of objects that exist independently of the infants' own sensory experience of them.

Why Object Is Not a Perceptual Representation

Perhaps the most studied topic regarding infant representational capacities is the concept object, in the sense of representations of substantial, three-dimensional, material bodies that exist independently of the observer. Are

Piaget and Quine correct that: (1) object representations are nonperceptual; (2) object representations are not available to young infants; and (3) object representations are built from sensory or perceptual primitives in the course of development?

Object representations, like depth representations, are clearly nonsensory, for they represent distal entities. Sensory representations capture object appearances such as color, retinally specified size and shape, and so on, but they do not represent objects as objects. Between them, Piaget and Quine offered several distinct reasons to consider object representations to be nonperceptual as well as nonsensory. First, Piaget argued that if perceptual representations are the output of modality-specific sensory analyzers, then object representations are not perceptual because they are multimodal. For adults, the representation of a visually perceived 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 along with Quine, believed that infants had to learn the cross-modal correspondences among the sensory 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.[2] Of course, learning contingencies among sensory representations in different sense modalities does not require nonsensory vocabulary. But the learning of such contingencies, Piaget thought, was the first step in transcending the initial sensori-motor primitives. Second—and here Piaget and Quine are also in agreement—if perceptual representations are limited to what currently experienced entities look like, feel like, taste like, and move like, objects cannot be represented as individuals that persist through time, independently of the observer. Quine agreed with Piaget that there would be no representations of permanent objects. As Quine pointed out, a perceptual vocabulary does not include fundamental quantificational devices. The child could not represent a given object as the same one as one seen earlier, for sensory representations do not provide criteria for numerical identity.

According to Quine, the infant endowed only with an innate perceptual quality space can sense similarity among experiences that are represented in this space (flesh-colored, milk-smelling experience at time 1; flesh-colored, milk-smelling experience at time 2), and the stable configurations of these qualities (color, shape, smell, sound) could certainly be learned. This would enable the child to recognize instances of mama-experience. In Quine's words, "his first learning [of the word 'mama'] is a matter of learning how much of what goes on around him counts as the mother" (1960, p. 92). But being able to recognize instances of mama-experience is not the same as representing one enduring mother —the same one today as yesterday.

Quine emphasized how much an ontology exhausted by a perceptual quality space differs from one articulated in terms of enduring individuals. For example, in one passage, he speculated that the baby reconceptualizes his mother once he has mastered the scheme of enduring and recurring physical objects. He also insisted that our adult commonsense ontology is a cultural construction, just as the concepts that articulate scientific theories are cultural constructions. Just as explicit theories embody their ontological commitments in language and formalisms, so too our commonsense ontology is captured in language. Indeed, the process Quine envisioned though which babies transcend the innate perceptual quality space and master the ontology of enduring and recurring physical objects crucially involves language acquisition. Quine proposed that 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 8 through 12 present a sympathetic characterization of Quinian bootstrapping processes, arguing for their role in the construction of new representational resources. My disagreement with Quine is straightforwardly empirical; in my view of conceptual development, he might be right. Rather, his picture of the infant just turns out to be false.

What Piaget wrote about under the rubric of "object permanence" comes to the same thing as what Quine wrote about as "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 isn't object permanence, but rather is some learned contingency, such as an appropriate instantiation of "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 sensori-motor, perceptual, and spatio-temporal 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's 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 spatio-temporal continuity (objects do not go into and out of existence). 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 on 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 what he called 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 (i.e., between ages 2:0 and 3:0). Therefore, evidence that 2- to 6-month-olds have these representational capacities challenges 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 two months of life. At the end of this chapter, I shall return to the question of innateness.

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 on the part of the behaving subject. 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 simply have 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. In 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 classic 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, multimodally specified, objects. With respect to intermodal correspondences, he observed infants being startled when they made a fast movement of their own hand across their visual field, and he assumed this meant that they did not know what their own hand looked like and that they could not relate a representation 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 he 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 sensory 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. Indeed, he discovered a second important phenomenon, the A-not-B error, which he took as knockdown evidence 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 that they were tracing the identity of the object through changes in location. It was not until 18 months or so, when infants could solve the hidden displacement problems, that Piaget (1954) was willing to credit them with a representation of object permanence. In the hidden displacement problems, an object is hidden by hand in location A, and the infant sees the closed hand move from A to B. The infant looks first in A and succeeds on this problem when, not finding the object at A, the child goes immediately to B. Piaget argued that this behavior requires representing the absent object and reasoning about its unseen movements.

These Piagetian observations are extremely reliable. They have been replicated replicated countless times, and were even incorporated into an infant "IQ test," because reaching these milestones markedly later than Piaget found sometimes reflects mental retardation (Bayley, 1969). Nonetheless, recent methodological advances have provided a wealth of data that reveal that Piaget underestimated the representational capacities of young infants.

Intermodal Representations

The empiricists believed that learning the intermodal correspondences between sensed properties of objects is the process through which representations of objects are built. The relevant correspondences include how visual appearance of texture is correlated with how that texture feels when touched, how visual appearance of shape is correlated with how that shape feels when touched, and so on, as well as correspondences between visually specified locations and the effects of reaching to proprioceptively specified locations. For the empiricists, there was nothing more to object representations than representation of such correspondences. Piaget disagreed with the empiricists on the issue of whether representations of objects can be cashed out in sensory and spatio-temporal vocabulary, but he agreed with them that intermodal representations such as those listed above had to be learned and that this learning is an essential part of the process through which nonperceptual 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 representations that go beyond sensory vocabulary—no representations of individuated, spatio-temporally continuous objects that exist independently of themselves. Second, there is now massive evidence that intermodal representations are innate and certainly not learned through the associative mechanisms Piaget and the empiricists imagined. Neonates orient visually to a location specified by a sound; neonates represent the correspondence between visually and tactually specified shapes; and neonates represent the correspondence between visually specified and proprioceptively specified facial gestures. Two experimental results can stand as examples from this large and convincing literature.

In the first example, Andrew Meltzoff and his colleagues allowed neonates to suck on a strangely shaped pacifier—either a smooth cube or a sphere with bumps all over it. These babies were only a few days old and had never had anything in their mouths other than nipples and their own hands. 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 (Meltzoff & Borton, 1979). A second example also comes from Meltzoff's laboratory. He and Keith Moore showed that neonates would imitate the facial gestures of an experimenter (mouth opening, tongue protrusion). Chapter 5 considers the significance of this result for the characterization of core cognition of human agents, but for now it is enough to show innate representations of the correspondence between what another's face looks like and the actions and feel of their own face (Meltzoff & Moore, 1977;

see Myowa-Yamakoshi, Tomonaga, Tanaka, & Matsuzawa, 2004, for a replication with an infant chimpanzee).

These data, along with those reviewed above ("A Historical Aside"), show that representations of people and objects, including their locations in space, are specified intermodally in neonates. Infants do not have to learn which sensations in one modality predict those in another—either in the service of learning the associations that the empiricists took to constitute depth representations nor the associations that they took to constitute object representations.

Criteria for Individuation and Numerical Identity of Objects; Object Permanence

My targets in this chapter are Piaget and Quine. My aim is to convince you that infants have an innate capacity to represent objects as existing independently from themselves and an innate capacity to quantify objects just as do adults. I will proceed in two steps: first, by reviewing the evidence that by 2 to 5 months of age infants have these capacities and then by turning to the question of innateness.

By 2 months of age, infants represent objects as spatio-temporally 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. The methods that show this were not available to Piaget. The literature I review in these early chapters draws on 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 highly improbable event. The robust result is that infants look longer at improbable or 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 versus nonelevated looking times as a source of data concerning the infant's representations of the ongoing events. In chapters

3 through 5 I consider further the nature of the infant's representations of these events. Here, I merely argue that infants' representations of these events are articulated in terms of a concept object, and they begin to characterize that concept in terms of evidence concerning the extension of the representation and its conceptual roles.

Renée Baillargeon, Elizabeth Spelke, and colleagues (1985) carried out the first violation-of-expectancy study that was brought to bear on infants' representations of objects as continuing to exist when out of sight. Four-month-old infants were habituated to a screen rotating 180°, as shown in Figure 2.1a. After habituation, an object was placed in the path of the screen on its downward trajectory, and one of two events ensued. In possible outcomes, the screen was rotated until it touched the object and then rotated back to its initial position (Figure 2.1b). In impossible outcomes, the screen was rotated through the space occupied by the object by the full 180° (Figure 2.1c). Infants looked longer at the

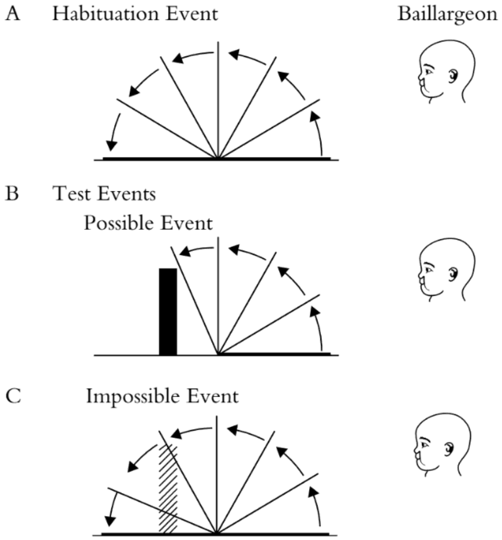


Figure 2.1. Diagram of Baillargeon, Wasserman, & Spelke (1985) rotating screen paradigm. a: habituation events. b: possible outcomes. c: impossible outcomes. Reprinted from Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in 5-month-old infants. *Cognition*, *20*, 191-208, with permission from Elsevier.

impossible outcome than at the possible one. Later studies revealed this pattern of results in infants as young as 2 months of age. These data were the first to suggest that very young infants represent an object placed behind a barrier to exist even when out of sight, as well as that infants' representations of object motion are constrained by the principle that one object cannot pass through the space occupied by another.

I more fully illustrate the logic of violation-of-expectancy studies with another design from Spelke's laboratory (Spelke, Kestenbaum, & Simon, 1995). This study was also aimed at exploring whether young infants represent objects as continuing to exist when out of sight. It is particularly relevant here for it raises the issue of how babies individuate objects and trace numerical identity over time, thus bearing specifically on Quine's claims concerning the quantificational capacities of prelinguistic human infants.

Figure 2.2a schematically depicts an event shown to infants in this typical violation-of-expectancy looking-time experiment. Two screens are placed on an empty stage, and two objects are brought out, in alternation, from the opposite sides of the screens and then returned behind them. The 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. In full habituation, infants watch these events until their interest in them declines by some preset ratio (e.g., to the point that their final looking at the event is half the level of their initial looking times). The question we want to pose to infants 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 spatio-temporally 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 (to adults) outcome of two objects, or the impossible outcome of just one object (thanks to a simple magic trick; one of the objects is surreptitiously removed through a trapdoor in the rear of the stage). Then the stage is cleared, the familiarization event repeated, and the other test outcome

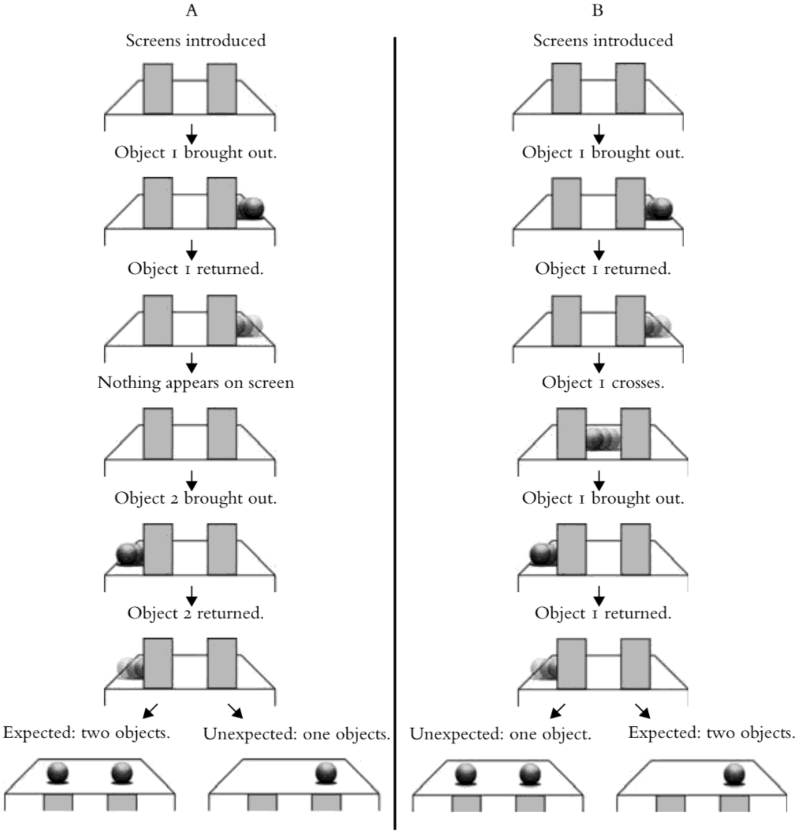


Figure 2.2. Diagram of the Spelke, Kestenbaum, & Simons (1995) split-screen spatio-temporal continuity paradigm. A: discontinuous motion condition. B: contiuous motion condition. Redrawn from Spelke, E. S., Kestenbaum, R., & Simons, D. J. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy, with permission from the *British Journal of Developmental Psychology*, *13*(2), 113–142. © The British Psychological Society.

revealed. Usually in these experiments there are three pairs of possible/impossible trials, alternating, with order counterbalanced across infants.

In these studies, infants look reliably longer at the impossible outcome of one object than at the expected outcome of two (see Figure 2.3a for typical data; Xu & Carey, 1996). These are the only actual data (the numbers) from an infant violation-of-expectancy looking-time study I will present in this book. (For most experiments, I present the pattern of data

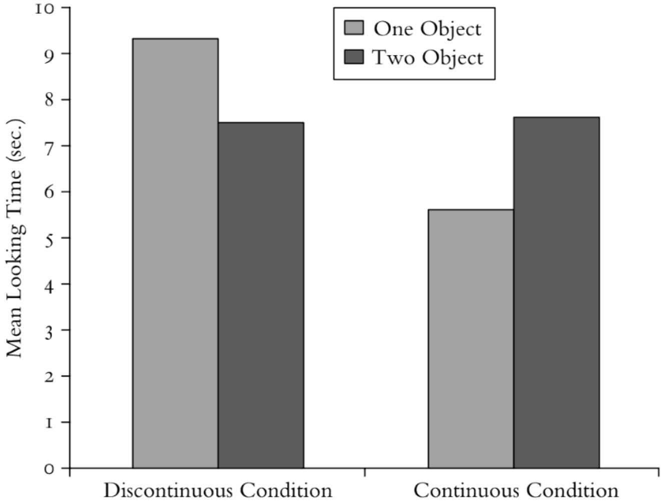


Figure 2.3. Ten-month-old infants' looking times during the test trials in the split-screen spatio-temporal continuity study (Xu and Carey, 1996). Reprinted from Xu, F., & Carey, S. (1996). Infants' metaphysics: the case of numerical identity. *Cognitive Psychology*, *30*, 111–153, with permission from Elsevier.

only qualitatively.) I present the numbers here to illustrate for those of you who have never read a technical paper using this method what exactly is meant when I say that infants looked longer at one type of outcome than at another (truth in advertising). Notice that the difference in looking times to the expected outcome (two objects) and the impossible outcome (one object) is small—less than 2 seconds, averaging across the 16 babies who were tested. Still, this is a very reliable result. Of 16 babies, 13 showed this pattern, and those who looked longer at the expected outcome showed a smaller difference in looking times between the two outcomes than did those who looked longer at the impossible outcome. Statistical analyses allow us to distinguish this pattern of data from random responding. Furthermore, this very pattern of data has been replicated many times. All of the looking-time studies I use for my arguments in this book have these characteristics: reliable and replicable differences in looking times between the expected and unexpected outcomes.

Of course, one must consider 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 might 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 hypothesis requires that infants distinguish arrays of one object from arrays of two objects, but it does not require that they represent the objects as continuing to exist behind the screen, nor that they use evidence regarding spatio-temporal 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 2.2b). 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 apparently prefer. 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 2.3b). 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 spatio-temporal discontinuity.

The original experiment using this design was carried out by Elizabeth Spelke and her colleagues with 4-month-old infants. Andréa Aguiar and Renée Baillargeon (1999, 2002), using essentially the same method, have shown that 2-month-old infants also expect objects to move on spatio-temporally continuous paths, even through occlusion.

Karen Wynn's (1992b) famous "addition/subtraction" experiments support the same conclusions: that infants use evidence of spatio-temporal discontinuity as a basis for individuating objects, and that they represent hidden objects as existing behind screens. 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. Her first study tested 5-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 2.4a. Infants watched as a single object was placed on an empty stage, and a screen was rotated up that hid the object. Then the infants watched as a hand brought in a second object and as the hand was withdrawn empty. The screen was then lowered, revealing either the expected outcome of two objects, or the unexpected outcome of one object. Looking times to outcomes of one and two objects in this condition were contrasted with those from the 2 – 1 = 2 or 1 condition (Figure 2.4b, in which case, the two-object outcome is unexpected and the one-object outcome is expected). Infants' patterns of looking were different in the two conditions; in the subtraction condition they looked reliably longer at the two-object outcome, whereas in the addition condition they did not. Wynn also found that infants succeeded in the 1 + 1 = 2 or 3 condition, looking longer at the unexpected outcome of three objects. Infants' attention is drawn whenever any number of objects other than precisely two is revealed after a 1 + 1 event.[3]

The implications of these results for our understanding of infants' representation of number will be explored in chapter 4; 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 from that one, and thereby take it to be 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 before they have learned the quantificational apparatus of their natural language; they distinguish one object seen on different occasions from two numerically distinct objects.

Aside: Why Do Infants Fail on Search Tasks?

Remember, Piaget's evidence that infants do not represent objects as continuing to exist in the absence of current sensory evidence of them

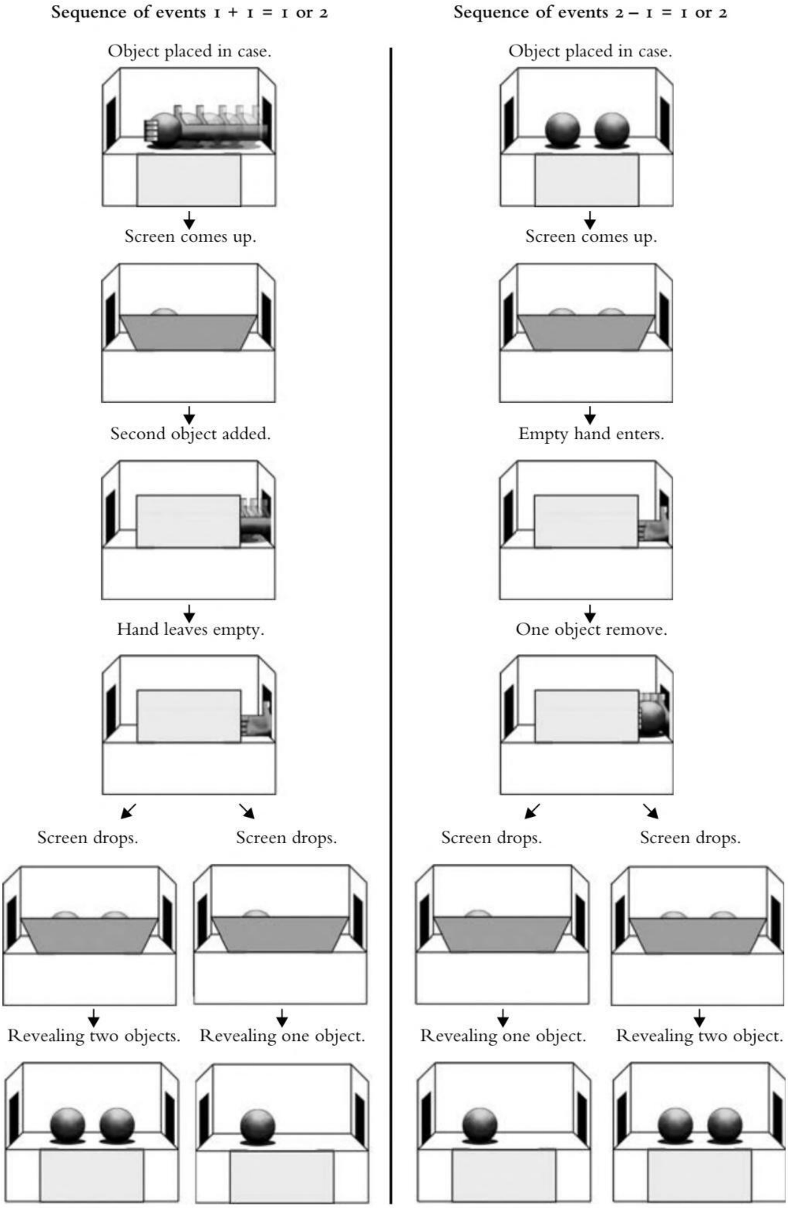


Figure 2.4. Diagram of Wynn (1992) addition/subtraction paradigm. 1 + 1 condition and 2 – 1 conditions. Redrawn from Wynn, K. (1992b). Addition and subtraction by human infants. *Nature*, *358*, 749–750, with permission from Macmillan Publishers Ltd.

was that they failed to retrieve them when hidden. But I have just reviewed the evidence that by 2 months of age, at least, infants represent objects as spatio-temporally continuous, tracking individual objects through space and time, even when occluded. Why then, do they fail in search tasks like the simple object-permanence tasks or the versions that reveal the A-not-B error?

There are two broad types of explanations for the failure on Piagetian tasks in the face of success on the looking-time tasks, and these are not mutually exclusive. First, any problem we set before an infant requires many different capacities. A failure on a given task may reflect the lack of some capacity other than the one that is the target of our interests. For example, many researchers 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 partly independent of the capacity to represent objects. The second explanation for earlier success on the looking-time tasks than on the reaching tasks begins with the observation that the capacity to represent some aspect of the world is not an all-or-nothing matter. Representations are graded in robustness or strength, are constructed in real time, and are subject to multiple interacting influences during the processes of construction (Munakata, McClelland, Johnson, & Siegler, 1997; Thelen, Schoner, Scheier, & Smith, 2001; Uller, Carey, Huntley-Fenner, & Klatt, 1999). These interacting influences guarantee that success on a wide range of tasks all putatively drawing on a common representational capacity will be task-dependent. 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 search differ in some respects from those that guide eye movement. For example, Yuko Munakata and her colleagues suggested that it is possible that more robust representations are required to support reaching than to evaluate consistency of visual models of the world.

Although it is easy to see how explanations like these might account for failures on some tasks in spite of the infants' having the representational capacity seemingly needed for those tasks, it is not easy to find evidence for any particular version. It is not impossible, though. For the sake of illustration, let's see how some of these ideas play out in understanding the developmental course of the A-not-B error, which wanes between ages 8 and 12 months or so.

Possible Explanation 1: Frontal Cortex Maturation

Adele Diamond and Patricia Goldman-Rakic offered an explanation of the first type—of the failure of young infants in the A-not-B task, in spite of the capacity to individuate and track objects through occlusion—in terms of a lack of a necessary prerequisite for task performance (Diamond, 1991; Diamond and Goldman-Rakic, 1989). Diamond and Goldman-Rakic began with the observation that the A-not-B task closely resembles a task used to diagnose frontal lobe 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 food in the immediately previous trials was hidden in the same well as in the current trial or the opposite one.

There is massive evidence for frontal lobe involvement in DR. Lesions in prefrontal cortex (specifically dorsolateral prefrontal cortex) of adult monkeys disrupt performance in DR tasks. 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 seconds. 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 seconds to 10 seconds 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, they disrupt so that performance falls to the level of 1.5-month-old infants.

Diamond and Goldman-Rakic suggested that the maturational change in prefrontal dorsolateral cortex taking place in infant *Rhesus* monkeys between ages 1.5 and 4 months occurs in infant humans between ages 7.5 and 11 months, and it at least partially underlies 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 solving the A-not-B task increased from 2 seconds at the youngest age to 10 seconds 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 they 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 seconds (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, she reasoned that if maturation of the structure underlies the parametric improvement on this task between ages 7 and 12 months of age, 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 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 she 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 the Plexiglas 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 an 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 Plexiglas box task, repeating the previously successful reach in the A-not-B task). Also, the 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. Diamond's work gives us evidence that the A-not-B error does not reflect a limit in the infant's representation of objects as spatio-temporally continuous, continuing to exist when occluded but, rather, reflects immature executive function that limits the means/end problem solving of infants under 1 year of age.

Possible Explanation 2: The Dynamic Systems Account

In a series of influential writings, Esther Thelen, Linda Smith, and their colleagues have discovered several new phenomena and have systematized the empirical literature concerning the A-not-B error (e.g., Smith, Thelen, Titzer, & McLin, 1999; Thelen et al., 2001). They argue that the error could arise from complex interactions among the multifaceted processes that enter into motor planning, processes that unfold over time. Thelen, Smith, and their colleagues stress that whether the infant makes the error or not is dependent on 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. Their model makes such novel predictions (which have been confirmed) as that the probability of the error will decrease if the child changes posture between successive trials!

In Thelen's model of motor planning, three distinct representations are built up over time, each having its own dynamics (rate of buildup, capacity for stability and self-maintenance, time course of decay), and they interact in a common motor workspace to create a plan to reach to A or B (or neither). The three distinct representations are (1) a representation of the task environment (that establishes the locations of A and B, and maintains them as distinct or as equally or differentially salient); (2) a representation of the cued location of a given trial; and (3) a representation of the previous movements, in which this representation is influenced by the entire history of movements, highly weighting the most recent ones. These representations are integrated in the process of planning a movement; a movement to A or to B ensues when a threshold of activity in the motor workspace 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 the developmental change between 8 months (A-not-B errors likely at delays greater than a few seconds) and 12 months (A-not-B errors unlikely, within a wide range of task parameters, at delays as long as 10 seconds) is modeled in terms of a change in a parameter called cooperativity. Cooperativity reflects the differentiation within motor space and the capacity for creating and maintaining a stable representation of the cued location.

Although Thelen and Smith's account differs from Diamond's in many respects, both place the A-not-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. Memory of the past action has a much longer time course of decay, the limits of which, as Thelen et al. point out, have not been systematically studied. If the processes that form and maintain the short-term memory of the cued location are fragile, an A-not-B error is thereby likely to occur. In sum, Thelen and Smith, on the one hand, and Diamond, on the other, agree that the A-not-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 in the first explanation. They sometimes deny the usefulness of the construct object representation or even representation at all. I find this puzzling. Their own model explicitly depends on three different types of representations: the task context, the cued location, and past acts. That these representations are formulated over motoric space, that they evolve over time, and that they interact in complex ways does not make them nonrepresentations.

Although these models of dynamic systems crucially depend on representations, the representations in this case are certainly sensori-motor ones. At other places, Thelen and her colleagues argue that it does not make sense to ask when infants "have" representations of objects. They claim that this is a badly mistaken question because representations are always manifested in behavior and thus their expression is always subject to the dynamic interaction of many different processes. This is undoubtedly true, and I will often rely on this fact in the pages to come (e.g., in explaining infants' failures on A-not-B tasks in spite of their capacity for object representations!). But this observation does not discharge the responsibility to account for the origin of the capacity to represent objects. Either this capacity is innate or it is built by some learning process, and such a learning process would necessarily occur at some particular point in time.

It is true that representations of objects play no role in Thelen and Smith models at all. The representations are representations of locations, with strengths determined by stimulus salience and dynamic factors. 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 and her colleagues discovered that infants will reach into the containers even when there are no objects in them, and that merely waving one of the visible lids, or touching it, would induce a reach into a particular one. That is, a wave or a touch would serve as a specific cue to a location on a particular trial (Smith et al., 1999).

It is not surprising that a model of the last stages of planning a reach is formulated in a motor workspace 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 —a particular object. 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-not-B error? I think not.

My colleagues and I have recently developed a search task that can be used to explore object representations in 10- to 12-month-old infants. Several studies using this method demonstrate that representations of objects guide the reaches of 10- to 12-month-old infants (Feigenson & Carey, 2003; Van de Walle, Carey, & Prevor, 2000). In this task, infants are introduced to a box into which they can reach but cannot see. We measure infants' search behavior as a function of what they have seen placed into the box. For example, on some occasions infants see two objects placed into the box and on other occasions they see only one object placed inside, after which the box is handed to the baby. There is always only one object in the box (the other having been surreptitiously removed on the two-object trials). The infant reaches in and retrieves the object, and the measurement period of interest is that which follows. Does the infant demonstrate, by his persistence of search, that he represents a second object inside it? Success on this task is longer searching on two-object trials because there should be a second object in the box than on one-object trials because the only object the child saw emerging from the box has been retrieved. Both 10- and 12-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 hidden 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.

Although I have criticized Thelen and Smith's arguments against mental representations in general, and their claims that it does not make sense to ask whether infants' "have" representations of objects, their models provide insight into why infants with the capacity to represent hidden objects make at the A-not-B error. Both Thelen and Smith's work and Diamond's provide detailed accounts of the complex processes involved in tasks that are used to characterize infants' representational abilities (in this case, in Piagetian search tasks). Variables that influence these processes can produce apparent failures on a given task even if the representational capacity in question is in place. Of course, positive evidence for that capacity is still required; in the present case, the positive evidence that young infants have the capacity to form representations with the content object derives from the looking-time studies reviewed above.

Are Object Representations Innate?

I have argued that very young infants represent objects as spatio-temporally persisting. 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 even when the child has no direct perceptual access to them. Quine's and Piaget's specific accounts of the origin of the capacity to form object representations cannot be right.

Still, the youngest age of participants in the violation-of-expectancy looking-time studies reviewed so far in this chapter is 2 months. Is it possible that younger babies' representations are formulated over sensory or perceptual primitives? Could the capacity to represent and quantify over objects displayed in these looking-time experiments be built between birth and 2 months of age? In the last pages of this chapter, I present arguments that convince me that perceptual input analyzers that yield representations of objects are most likely innate.

This question is particularly trenchant because there is one piece of evidence from a looking-time paradigm that suggests that the capacity to compute object representations is 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 2.5a). 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 are the processes that result in object individuation. Under what circumstances, if any, does the infant establish a representation of a single, spatio-temporally continuous (i.e., connected throughout) object extending behind the barrier, rather than two numerically distinct objects? Philip Kellman and Elizabeth Spelke used the violation-of-expectancy looking-time method to answer this question (Kellman & Spelke, 1983). 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 2.5b) is revealed than if a continuous rod (Figure 2.5c) is revealed. Building on this work, Scott Johnson and Richard Aslin have shown

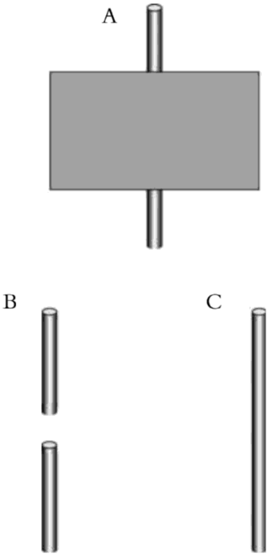


Figure 2.5. Diagram of stimuli for the Kellman & Spelke (1983) broken rod experiments. a: habituation stimulus (rod moves back and forth behind the screen). b: two rod outcome. c: single-rod outcome. Redrawn from Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*(4), 483–524, with permission from Elsevier.

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 (Aslin & Johnson, 1996; Johnson & Aslin, 1995).

Newborns, however, are different. Alan Slater and his colleagues found that newborn infants display the opposite pattern of looking times (Slater, Morison, Somers, Mattock, & Brown, 1990). Habituated to the array in Figure 2.5a, they look longer at the completed rod (Figure 2.5c) than at the broken rod (Figure 2.5b), as if the former were a novel stimulus for them. Slater's findings have been taken to show that between birth and 2 months of age, infants learn that common motion of two visible portions of objects protruding from behind a barrier is likely to be part of one and the same object.

There are, however, alternative explanations for the neonate's failure in the face of 2-month-olds' success, other than that the processes that create object representations are constructed through learning in the first two months of life). Just as Diamond argued in the case of developmental changes in the A-not-B error, it is possible that maturation of capacities other than those that create object representations per se underlie the change between newborns and 2-month-olds. Alternatively, it is possible that neonates need more redundant information still, compared to 2-month-olds, just as 2-month-olds do compared to 4-month-olds, for amodal completion, and that 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.

How might we decide between an explanation of the developmental change that involves learning that a single object is likely to be found behind the barrier and one that involves developmental changes in processes that are inputs to an innate computational device? Three empirical considerations lead me to favor the nativist view that the capacity for amodal completion is the product of evolution and it does not have to be constructed through learning processes.

First, it is not hard to imagine ancillary capacities that might await development before infants can succeed at completing the rod behind the barrier. They must notice the correlated motion of the two ends of the rod: this is the input to the computation that creates a representation of the single rod. Young infants have a notorious difficulty deploying their attention. Two sources of data suggest that one problem faced by very tiny babies is just this failure to notice the correlated motion: Two-month-olds are less likely to complete the rod behind the barrier if the barrier is wider, and increasing width plausibly makes it more difficult to notice the common motion (Condry, Smith, & Spelke, 2001). Confirming the necessity of doing so, eye-tracking studies show that 3-month-olds complete the rod only if they scan between the two ends of the rod during familiarization (Amso, Davidson, & Johnson, 2005). The Amso et al. study showed that young infants' attention is captured by the motion of one portion of the rod along the edge that specifies the occluder.

Second, even stronger than evidence consistent with some possible way of explaining away a failure is positive evidence that neonates have the capacity. A recent study of neonates presented the stimuli stroboscopically, showing the end points of the movement only and thus removing interference from the encoding of relative motion along the edges. The neonates generalized habituation to the complete rod, just as do 2-month-old and older infants (Valenza, Gava, Leo, & Simeon, 2004). Thus, amodal completion appears to be innate in humans.

Finally (and this is indirect evidence), the capacity for amodal completion is innate in chickens (Regolin & Vollortigara, 1995). Neonate chicks imprinted on a red triangle, partially hidden behind a barrier, huddle next to a completed triangle, rather than on a broken one, the first time the barrier is removed (Figure 2.6). These newborn chicks had no opportunity to learn what stimulus conditions predict a complete figure as opposed to a broken one under these circumstances. Indeed, even the spatio-temporal continuity implicated in object permanence is the output of innate perceptual analyzers in chicks. Newborn chicks, imprinted on a ball, which have never in their lives seen any object go behind a barrier (and thus could not have learned about spatio-temporal continuity), search behind a screen for the ball the first time it disappears there. They even avoid the A-not-B error! Of course, that object permanence is innate in baby chickens does not mean it is innate in human babies. Nonetheless, these studies provide an existence proof that it is possible for the capacity to represent objects as spatio-temporally continuous, even under conditions of occlusion, to be manifest without learning.

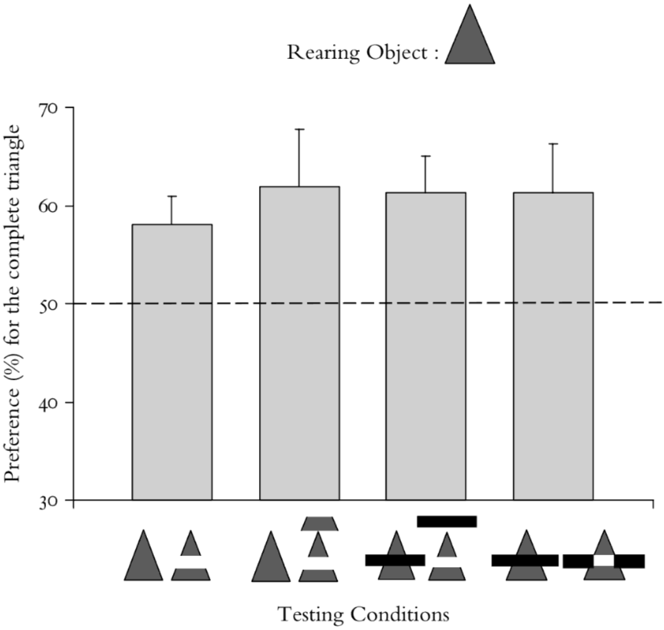


Figure 2.6. Preference for the complete (or amodally complete) stimulus of each comparison pair. Group means +/- standard error depicted. From Regolin & Vallortigara, 1995. Regolin, L., & Vallortigara, G. (1995). Perception of partially occluded objects by young chicks. *Perception & Psychophysics*, 57(7), 971–976. Reprinted with permission from Psychonomic Society, Inc.

Even if we did not have these empirical results in hand, other considerations would bear on deciding between the learning account of the change in performance between 0 and 2 months of age and the alternatives. Those who favor a learning account of the change between birth and 2 months of age need to sketch one. Quine's linguistic bootstrapping process was his answer to this question, and that hypothesis is already ruled out by the existence of object representations and the capacity for divided reference in clearly prelinguistic infants. 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 see how infants could learn statistical regularities stated over perceptual and spatio-temporal primitives—noting that certain patterns of occlusion predict certain patterns of reappearance, for example, or that certain patterns of common motion predict spatio-temporal continuity of the elements that are moving together. Statistical analyses—for example, of the sort so well modeled in connectionist architectures—could accomplish such learning; and indeed, there are successful models that do just that (e.g., Munakata et al., 1997; Mareschal, Plunkett, & Harris, 1999). However, these generalizations are not stated over object representations. Furthermore, even if they were, they would not constitute representations of object permanence unless the system represents the object as the same one that went behind the barrier. As Gary Marcus (2001) points out, either the current simulations cannot do so or they build in this capacity from the beginning, thus accomplishing interesting learning, but not the learning of spatio-temporal continuity itself . Similarly, even if generalizations about common motion and connected, filled spatial regions were learned, 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. If it is true that object representations cannot be expressed in a sensori-motor or perceptual vocabulary, there is a serious learnability issue of how they could be learned from statistical generalizations over that vocabulary.

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 in principle a 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 nestling indigo buntings. A proposal for a 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 even a successful proposal for a plausible learning mechanism would not