WHAT BABIES KNOW

VOLUME 1

CORE KNOWLEDGE AND COMPOSITION

ELIZABETH S. SPELKE

OXFORD SERIES IN COGNITIVE DEVELOPMENT

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

Library of Congress Cataloging-in-Publication Data
Names: Spelke, Elizabeth, 1949– author.
Title: What babies know : core knowledge and composition :
Volume 1 / Elizabeth S. Spelke.
Description: 1 Edition. | New York : Oxford Universsity Press, 2022. |
Series: Oxford cognitive development series |
Includes bibliographical references and index. | Contents: v. 1.
Identifiers: LCCN 2022000245 (print) | LCCN 2022000246 (ebook) |
ISBN 9780190618247 (v. 1; hardback) | ISBN 9780190618278 (v. 1; epub) |
ISBN 9780190618261 (UPDF) | ISBN 9780190618254 (Digital Online)
Subjects: LCSH: Infants—Development. | Child development.
Classification: LCC HQ774 .S64 2022 (print) | LCC HQ774 (ebook) |
DDC 305.232—dc23/eng/20220118
LC record available at https://lccn.loc.gov/2022000245
LC ebook record available at https://lccn.loc.gov/2022000246

DOI: 10.1093/oso/9780190618247.001.0001

1 3 5 7 9 8 6 4 2

Printed by Integrated Books International, United States of America

Contents

Acknowledgments

When a book covers 50 years of research and takes 20 years to write, its author will have a lot of people to thank. First are my teachers. Jerome Kagan's boundless enthusiasm for developmental psychology got me started. Eleanor Gibson's vision of perception and exploration and her extraordinary talents as an experimenter set me on the path that all my research on infants has followed. Ulric Neisser's broad vision of a cognitive psychology in which basic and applied research go hand in hand is a continuing inspiration, as is Stephen Emlen's passion for behavioral ecology and his talent for revealing the remarkable cognitive feats of apparently simple animals.

Second are the generous colleagues at four institutions who have helped and challenged me along the way. At the University of Pennsylvania in the late 1970s and early 1980s, in a psychology department that was made in heaven, I am especially indebted to Jonathan Baron, Randy Gallistel, Rochel Gelman, Henry Gleitman, Lila Gleitman, Jacob Nachmias, Daniel Osherson, David Premack, and Paul Rozin. At Cornell in the late 1980s and early 1990s, my greatest debts are to Frank Keil and Carol Krumhansl; at MIT in the late 1990s, they are to Nancy Kanwisher, Earl Miller, Steven Pinker, Mary Potter, Pawan Sinha, Mriganka Sur, Kenneth Wexler, and Patrick Winston. I am now blessed by the extraordinary cognitive science community that has spanned Harvard and MIT over the last two decades, especially Mahzarin Banaji, Susan Carey, Patrick Cavanagh, Alfonso Caramazza, Edward Gibson, Marc Hauser, Nancy Kanwisher, Stephen Kosslyn, Roger Levy, L. Mahadevan, Ken Nakayama, Steven Pinker, Rebecca Saxe, Laura Schulz, Jesse Snedeker, Joshua Tenenbaum, and Tomer Ullman.

Other colleagues have hosted me on sabbaticals and extended or repeated gatherings where my thinking has taken shape. My first sabbatical at MIT (1982–1983) introduced me to the field of cognitive science, thanks to the patience and generosity of Ned Block, Sylvain Bromberger, Susan Carey, Noam Chomsky, Jerry Fodor, Merrill Garrett, Ken Hale, Dan Osherson, Steve Pinker, Whitman Richards, Shimon Ullman, Ken Wexler, Patrick Winston, and especially another lucky visitor, Jacques Mehler. I owe my second sabbatical, at the Ecole des Hautes Etudes en Sciences Sociales (1984–1985), to Jacques, and also to Bénédicte de Boysson-Bardies, François Bresson, Celia Jakubowicz, Roger Lécuyer, Marie-Germaine Pêcheux, Gilberte Piéraut-le Bonniec, Scania de Schonen, and especially Arlette Streri. For later sabbaticals in France, I thank Emmanuel Dupoux, Dan Sperber, Pierre Jacob, Daniel Andler, Jacqueline Fagard, Arlette Streri,

Kevin O'Regan, Henriette Bloch, Stanislas Dehaene, Gretty Mirdal, Simon Luck, Saadi Lahlou, and their institutions: the Institut Jean Nicod (2000–2001), the Laboratoire de Sciences Cognitives et Psycholinguistique (2000–2001), the Cognitive Neuroimaging Unit at the Service Hospitalier Frederic Joliot (2000–2001), the Laboratoire de Psychologie du Développement et de l'Education de l'Enfant (2006), the Laboratoire de Psychologie de la Perception (2006), and the Paris Institute of Advanced Studies the College de France, and Neurospin (2017 and 2022). I also thank the French CNRS, the Guggenheim Foundation, and the James McKeen Cattell Fund for support during these sabbaticals.

Like many cognitive and brain scientists, I am indebted to Michael Gazzaniga and the McDonnell Foundation for founding and sustaining the memorable Summer Institutes in Cognitive Neuroscience. Those institutes are gone but have many descendants: I am grateful to Alfonso Caramazza and Giorgio Vallortigara for creating a space, at the Center for Mind-Brain Sciences in Rovereto, where research on infants and animals synergizes and thrives, and to Tomaso Poggio, NSF, and the marvelous MIT Center for Brains, Minds, and Machines, where research on cognition in infancy, together with research on many other topics, approached from multiple angles, has found a home. I am also indebted to Mike Gazzaniga again, and to the Dana Foundation, for helping me to realize that developing human minds care about the arts as well as the sciences.

Fortunately, many individuals and institutions have fostered interdisciplinary research in the developmental cognitive sciences through lively and regular research gatherings. I am especially grateful to John Morton, Annette Karmiloff-Smith, Mark Johnson, and the much-missed Cognitive Development Unit in London; Claes von Hofsten, Kerstin Rosander, Gustaf Gredebäck, and their groups at the universities of Umea and Uppsala; Gyorgy Gergely, Gergely Csibra, and their group at the Central European University in Budapest and Vienna; Stanislas Dehaene, Antonio Battro and their meetings at the Pontifical Academy of Sciences; Willem Levelt, Peter Hagoort, and Melissa Bowerman at the Max Planck Institute for Psycholinguistics in Nijmegen; Pierre Pica, Tomer Ullman, and the Lorentz Center in Leiden; the Rockefeller Institute in Bellagio, and especially John Bruer, Susan Fitzpatrick, Kathy Hirsh-Pasek, Juan Valle Lisboa, Alejandro Maiche, Marcela Pena, Sidarta Ribeiro, and Mariano Sigman, the guiding lights behind the essential and irreplaceable Latin American School for Education, Cognitive, and Neural Sciences. Special thanks go to the Jean Nicod Prize committee and the Nijmegen Lectures committee: The core ideas in this book came together in their two series of lectures and in the discussions that followed.

Education, like science, never ends. Generations of younger scientists, who were and are extremely busy doing wonderful work, have brought me new ideas, methods, and findings, while letting me join some of their efforts. Five of them have profoundly influenced my research and thinking: Stanislas Dehaene,

Ghislaine Dehaene-Lambertz, Esther Duflo, Nancy Kanwisher, and Joshua Tenenbaum.

For 45 years, students, postdocs, and lab managers have been the true drivers of my research. Those who focused on infants appear throughout this book; many others will appear in its successor. Two individuals deserve special mention. Philip Kellman built my first independent baby lab, with material aid from Rochel Gelman and the Penn Psychology Department. And Kirsten Condry rebuilt the lab twice, at three institutions, while serving as its heart, soul, and central nervous system. In 1996, she ran our Cornell lab for its last year while overseeing the design of our new lab at MIT. Five years later, she designed the beautiful labs where Susan Carey and I now work at Harvard, and she was central to the launching of our new program, with Jesse Snekeder, in developmental cognitive science.

Then there are the colleagues and friends who have sustained me throughout the period over which this book was conceived and written. Some of their names have already appeared, but they get a second thanks for making science and life so much fun: Mahzarin Banaji, R. Bhaskar, Ned Block, Susan Carey, Stan Dehaene, Ghislaine Dehaene-Lambertz, Judy Deloache, Carol Dweck, Jacqueline Fagarad and Remi Fagard, Ken Forbus, Randy Gallistel, Rochel Gelman, Dedre Gentner, Henry Gleitman and Lila Gleitman, Susan Goldin-Meadow, David Goldman, Peter Jusczyk, Nancy Kanwisher, Rachel Keen, Katherine Kinzler, Richard Kittredge, Tanya Korelsky, Barbara Landau, Susan Levine, Alejandro Maiche, Ellen Markman, Bill Meadow, Jacques Mehler, Yuko Munakata, Elissa Newport, Randy O'Reilly, Philippe Rochat and Rana Rochat, Kerstin Rosander, John Rubin, Kristin Shutts, Ted Supalla, Arlette Streri, Josh Tenenbaum, Claes von Hofsten, and Sandy Waxman. Our gatherings have softened the boundaries between labs and homes, cultures and generations, and teachers and students. There are reasons that these boundaries exist, but there also are benefits to entering a space where one no longer knows when one is teaching or learning, guiding or following, and working or playing.

Many people have read parts of this book over the years, giving me invaluable criticism, suggestions, and advice. Thanks to David Barner, Paul Bloom, Susan Carey, Alexandre Duval, Susan Gelman, Lila Gleitman, Susan Goldin-Meadow, Yuval Hart, Zoe Jenkin, Rachel Keen, Sang Ah Lee, Shari Liu, Lindsey Powell, Peng Qian, Brian Reilly, Josh Rule, Kristin Shutts, Amy Skerry, Tomer Ullman, and Brandon Woo. At Oxford University Press, Gwen Colvin and Martha Ramsey improved its style, and Martin Baum smoothly shepherded it through the publication process. Three outsiders performed the greatest acts of magic at the beginning and end of this process. Kara Weisman, between college and graduate school, was brave enough to work on this book in its earliest stages, when it was an utter mess, and designed the first prototypes of its figures. In its last

stages, Misha Oraa Ali created marvelous figures in record time, where they were most needed.. And Brian Reilly, between his teaching of French language, literature, culture, science, and medieval texts, intrepidly wrangled with the final version and miraculously emerged with its index. None of these people are to blame for the errors that remain.

This book is dedicated to Mae Bridget Spelke and to Joseph Alan Spelke Blass. As children and adolescents, they were part of all the sabbaticals and much of the travel and gatherings described here. They helped with experiments and played with generations of students and academic friends. They have also read and commented on the passages of this book that have been hardest for me to write. Neither has become a developmental cognitive scientist, but both have developed a passion for research that contributes to knowledge and aims to make the world a better place.

I've saved the most important acknowledgements for last. Not a word of this book could have been written without the families who have participated in the research it describes. Without parents who trust in science and are willing to contribute to it, experiments on infants would not exist. Moreover, I view the infant participants in this research as the book's true authors and my most valued collaborators and challengers. Special thanks to two sets of families: the infants and parents who participated in my first experiments, patiently putting up with my errors and false starts, and those who were willing, over the past two years and under the stresses of the pandemic, to spend time on Zoom, contributing to basic research on infant minds.

Prologue

How do we grasp abstract concepts like *circle, six, wish,* or *good*? What is special about human cognition? With perceptions and actions so similar to those of other animals, why do we alone develop new systems of knowledge, like astrophysics and medicine, and new technologies that remake the world? What is universal about human cognition? Beneath the variable knowledge and skills that support our diverse languages, cultures, religions, ideologies, and passions, is there a bedrock of assumptions, beliefs, and values that we all share? In this book and its sequel, *How Children Learn*, I aim to shed light on these questions by focusing on two others. First, what do human infants know at the time when their learning begins? And second, how do infants and children learn about the particular places, things, people, and events they encounter, and what makes their learning go so well?

Questions concerning the nature and sources of our abstract concepts have a long history, because such concepts present a puzzle. Many of them are so simple that preschool children talk about them, and so important that they stand at the foundations of a host of fundamental cultural achievements, including mathematics, technology, ethics, and the arts, but the concepts themselves are elusive. A perfect circle has no thickness and so cannot be drawn or touched. Six, a natural number, belongs to infinitely many sets with surprising properties: How can there be as many even numbers as integers, for example? Wishes are mental states that both are and are not part of the material world. And good knives, novels, liars, and deeds have little in common. How do we arrive at these concepts, given the limits to our experience? Neither we, nor the ablest machine we can build, will ever see a perfect circle, count to infinity, or touch a person's thoughts.

The uses we make of our abstract concepts reveal a striking feature of human cognition: We likely are the only animals who create new systems of knowledge over our cultural history and learn them over our lifetimes. Yet our minds and brains are so similar to those of other animals that much of our knowledge of our own capacities for perception, learning, memory, and action comes from research on other species: from the pioneering studies of Eleanor Gibson on visual space perception in newborn goats, and of David Hubel and Torsten Wiesel on the organization and development of visual cortex in cats and monkeys (chapter 1), to the landmark research of Edward Chase Tolman and John O'Keefe on the cognitive and brain processes supporting navigation and spatial memory in rats (chapter 3). These observations suggest a more focused version of my

second question: What are the distinctive qualities of our minds that allow us to use the experiences and neural systems that we largely share with other animals to develop new concepts and beliefs that are so different from theirs?

Our capacity to master new knowledge systems creates diversity within our species: People who live in different cultures, or who have lived at different times, have widely differing concepts, beliefs, skills, interests, and opinions. Consider, for example, how much attitudes toward child labor, capital punishment, homosexuality, or the role of women in public life have changed over just the last century, and how variable, across people, some of these attitudes are today. In the face of this variability, my third question also can be rephrased: Are there core cognitive capacities that stand at the foundations of human life in all cultures, and that allow a newborn infant to learn the language, concepts, values, and skills that structure life in the society she finds herself in?

That question brings me to the topic of this book and its successor. Human infants and young children face a formidable learning challenge. Equipped only with the universal capacities of our species, they must master all the commonsense knowledge required for life in the society and culture into which they were born. Strikingly, children accomplish a good part of this task without being taught. Preschool children learn their language, develop a commonsense understanding of how the world works, and take on many of the beliefs and values of the people in their culture before they enter school. Infants begin to learn these things before they begin to speak, even in cultures in which adults rarely speak to them. Infants and children learn not only in families with rich adult-child interactions but also in communities in which young children spend most of their time with peers. Even in cultures like ours, where parents widely believe that their children should be stimulated and instructed as well as nurtured and loved, infants learn all sorts of things that adults do not intend to teach them.

Of course, children are not the only adaptive learners. Many animals show exquisitely rapid and effective learning in biologically significant domains. Chicks, ducks, and geese learn to identify their mother through a rapid process of imprinting; birds learn to migrate over long distances from their winter grounds to the summer habitat in which they were born, building mental maps of the movements of the stars and of the local terrain at their birthplace; and rats learn, in a single, unpleasant trial, to avoid a poisonous food. These learning processes, however, are not flexible: The mechanisms by which chicks learn to identify Mom do not serve to identify the path home or the poisonous plant that abuts it.

Animals also have a remarkably general ability to learn whatever contingencies their local environment presents, even when those contingencies are arbitrary, like the bell rung by Pavlov to announce the arrival of food for his dogs. This general learning process, however, is slow, as it depends on the gradual accumulation of information that one event reliably heralds another. In recent years,

the field of artificial intelligence has produced machines that speed up some of the slow, general learning processes found in humans and animals. As a result, the capabilities of the biggest and fastest machines now exceed those of humans in domains like chess and Go, by processing far more information than any person could accumulate in a lifetime. In contrast both to animals and to these machines, however, human infants and preschool children gain a commonsense understanding of their environment through learning processes that require far less information, and that are both fast and flexible. How do they do this?

This book presents my best attempt to answer a piece of this question, focusing only on children's knowledge and learning in infancy. My answer comes in two parts. First, infants' learning rests on a set of cognitive systems that we share with animals and that evolved over hundreds of millions of years. At least six distinct systems serve to represent highly abstract properties of the unchanging navigable environment, of movable objects, of number, and of the living, animate, and social beings who populate our world. The systems share a constellation of properties and limits that distinguishes them from the cognitive systems that philosophers and psychologists have traditionally recognized: perceptual systems, action systems, and belief systems. I call them systems of *core knowledge*.

Second, humans have evolved one set of cognitive capacities that are unique to us: capacities to learn a natural language and to use that language for thinking, for communicating, and for grasping the thoughts of others. During the year this book covers, infants do not use language for communication, but they are continually engaged in learning their native language. The language that infants learn, beginning in the womb and ending just before most of them start to speak intelligibly, allows them to compose new concepts from the concepts of core knowledge. And the language infants hear, from the people in their social world who speak to one another (and in many cultures, to them), provides guideposts for organizing and using these new concepts.

By the end of the infancy period, children have gained the basic tools they need to develop the commonsense knowledge that life in their culture requires. In *How Children Learn*, I will ask how children accomplish this task, beginning at 1 year of age and continuing over the years that separate infancy from the onset of formal schooling. That book will focus on children's developing knowledge of object forms and functions—knowledge that underlies our prolific invention and mastery of tools; of spatial symbols including pictures, maps, and the alphanumeric characters supporting reading and calculation; of the natural numbers; of Euclidean geometry; of mental states as propositional attitudes; and of the structure of the social world. This book ends where the second will begin. It focuses primarily on cognitive development over the course of the first year: a time when infants' learning is propelled by a penchant for observing, exploring,

experimenting, and engaging with others, guided first and foremost by their systems of core knowledge.

Core Knowledge

Drawing on more than 40 years of research, this book introduces six core systems (figure P.1).[1] The most richly studied core system focuses on *places* in the persisting, navigable surface layout. It provides us with a sense of where we are, supports the construction of mental maps of the terrain through which we move, and anchors our memory for the events we experience. Another core system focuses on *objects:* nonliving bodies and their motions. It is the foundation for our commonsense understanding of the physical world. A third system focuses on *number* and represents the approximate numerical magnitudes of sets of objects or events. Among other functions, it supports learning about the statistical properties of the things and events we experience. Core knowledge of places, objects, and number has been widely studied in animals as well as humans, using the methods and perspectives of diverse disciplines in the cognitive, brain, and computational sciences, including experimental psychology, systems and cognitive neuroscience, and artificial intelligence and robotics. These are the simplest, best understood core cognitive systems, so I begin the book with them.

Using the methods and findings from studies of these systems, I hypothesize that infants have three more systems of core knowledge. The *form* system evolved, I believe, to represent the forms and functions of living beings that grow, provide our food, and defend themselves against us. For human groups living in close contact with nature, this system gives rise to commonsense knowledge of botany and ecology. For children and adults in industrialized societies, it is diverted to support learning of the forms and functions of artifact objects: learning at the foundations of tool use and technology. I have given it a name that applies in both these contexts. The *agent* system focuses on animate beings, including people, who act on objects and cause changes in them. It underlies our action understanding, action planning, causal reasoning, and grasp of people's intentions. The last system focuses on *social beings* who engage with one another, share their experiences, and form enduring bonds. For our species, this system supports learning about individual people and the network of social relationships that connects them to one another and to us. The agent and social systems together support children's learning about people and their mental states, and they form the core of our commonsense understanding of human societies and ethics.

[1] I thank Shari Liu for creating this figure.

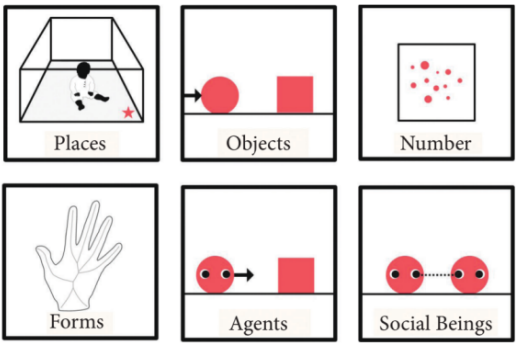

Figure P.1 Six domains of core knowledge.

Core knowledge systems have useful properties. First, each system captures a distinct, interconnected set of abstract concepts. For example, the place system represents the geometric structure of the navigable terrain over which animals move and objects reside, and the traversible paths that connect individual places to one another, and the social system represents the people in one's social world as having experiences like one's own that they share in states of engagement. I believe the core systems provide the foundations for the abstract concepts at the center of all our explicit knowledge. Second, core knowledge functions throughout life. Because each core system centers on abstract concepts, the systems capture fundamental, generalizable properties of the world, like the continuity of time and motion, the geometric properties of space, and the causal powers of animate beings. These systems give rise to our deepest intuitions, as adults, about the nature of the world and ourselves. The core systems, therefore, provide a common ground for communication between people in different cultures, with different beliefs and values.

Where relevant evidence is available, the core systems have been found to be innate, in a straightforward and unproblematic sense of this much-maligned term: They are present and functional on an infant's first informative encounters with the entities that they serve to represent. As a result, these systems support acting, exploring, and learning as soon as the need arises. In humans, moreover, the core systems support the development and use of new systems of knowledge. Although the core number system lacks the power and scope of the natural numbers, for example, children use it, together with other systems, to construct the natural number concepts at the foundations of mathematics. Core systems are more than scaffolds, however: Adults whose core number system is impaired by injury no longer can calculate with natural numbers as they did before.

Core knowledge systems have further properties. First and foremost, each system singles out entities in a distinct cognitive domain: The place system, for

example, represents the continuous, unchanging, extended surface layout, but not the objects within it. Similarly, the object system represents inanimate, cohesive, solid, and persisting bodies that move only on contact, but not animals, who have the power to generate their own motion, to perceive things at a distance, and to direct their actions to distant objects. The behavior of animals cannot be fully captured by the rules of contact mechanics in the domain of the object system. Together, the six systems carve our cognitive territory into more manageable units.

Second, the core systems are astonishingly ancient: They have been shaped by hundreds of millions of years of cognitive evolution. Some core systems are shared by animals as remotely related to us as fish, and aspects of these systems are shared by flies and worms. These ancient systems, found even in animals who lack a cerebral cortex, give rise to mental representations that are deeply inaccessible to our human, conscious minds. We can discover them through experiments, but not by introspection.

The ancient origins of the core systems limit the properties they serve to represent. Core systems capture deeply important, general properties of the entities in their domain, but they mostly fail to capture what those entities look like or how, in detail, they behave. To give just two examples, the place system represents the ridges, hills, valleys, and impassable borders of the ground surface over which we and other animals navigate, but it doesn't represent the walls of the infant's house, the objects that furnish it, or even the colors and patterns on the floors on which people walk. The youngest child navigators, eager to find their favorite toy that has been hidden in a room, will retrieve the toy by detecting highly subtle spatial perturbations on the room's floor but will fail to retrieve it by noting its proximity to a brightly colored wall or a salient visual pattern. Similarly, the social system represents the direction of gaze of a face with striking precision but fails to specify the physical characteristics that distinguish faces of the infant's species from those of other species. As a result, newborn human infants look longer at a still face whose eyes are looking at them than at an otherwise identical face looking elsewhere, but they are equally attracted by the direct gaze of a person, a sheep, or a bird.

These ancient core systems have a further property: They are modular, in all the respects described by the philosopher and pioneering cognitive scientist Jerry Fodor in his *Modularity of Mind.* In particular, core knowledge systems operate automatically, regardless of our intentions, thoughts, and beliefs, although their operation does require attention to entities or events that lie within their domain. Finally, human infants and nonhuman animals of all ages fail to combine the outputs of the core knowledge systems flexibly and productively. Neither young infants nor nonhuman animals are capable of fast, flexible learning when

faced with tasks that require new combinations of information from different core domains.

If one considers all these properties together, it becomes clear that core knowledge systems do not fit within the categories that appear in traditional treatments of human psychology. They are not sensory or perceptual systems, because they serve to represent abstract properties and relations. They also are not the central cognitive systems that underlie our explicit beliefs, decisions, and reasoning. For example, our actions and stated beliefs do not abide by the core principle that inanimate objects only change their motion on contact with other objects. Indeed, we believe otherwise, and we act on our explicit beliefs when we explain to our children how the earth moves around the sun. Core knowledge systems occupy a middle ground between perceptual systems and belief systems. Cognitive science needs this middle ground, I believe, to understand both human and animal minds.

Although core knowledge exists and functions in the minds and brains of animals and adults (and I will draw heavily on studies of those populations to elucidate its properties), I focus on human infants from birth to 12 months, for two reasons. First, studies of infants shed light on the contents and functioning of human minds before they are filled with culture-specific beliefs and attitudes. Infants are hard to study, because we have no intuitive access to what they know and they have limited means for expressing their knowledge. Nevertheless, experiments on infants provide the most direct access to the earliest emerging cognitive capacities at the foundations of our knowledge. Second, infants' learning lays the groundwork for a constellation of remarkable changes that occur toward the end of the first year: Infants begin to learn not only from others' actions but from their words, and they begin to communicate with others through gestures and patterns of shared attention to objects and events. One-year-old infants have begun to view people, and the objects that people talk about, in new ways.

These changes usher in a period of learning that has no parallel in the living world: the period to be covered in *How Children Learn*. Throughout this period, however, hidden but ever-present systems of core knowledge provide the central abstract concepts that children build on in later years. By using young infants' exceedingly limited behavioral repertoire for all that it is worth, developmental cognitive scientists can discover these concepts and trace the earliest learning they support. Young infants' minds reside in cognitive territory, between perception and belief, that is opaque to our mature, conscious experience. But as adults, our unconscious minds also inhabit that cognitive territory, and so the seeds that we discover in infants, through experiments using the multidisciplinary methods of cognitive science, bring us insight into ourselves.

Organization

This book has 10 chapters. The first chapter isn't about core knowledge; it focuses on infants' visual perception of surfaces and depth. I begin with this topic because research on perceptual development in infancy provided the primary methods by which cognitive scientists have discovered what infants know. Chapter 1 centers on the work of two extraordinary psychologists, Eleanor J. Gibson and Richard Held, whose work from the 1950s to the 1980s tackled ancient questions concerning the nature and origins of perception of the visible spatial layout. From distinct theoretical perspectives, but with converging methods, Held and Gibson turned a long-standing philosophical and scientific debate into a set of empirical questions that they, and other experimental psychologists, proceeded to answer. In doing so, they solved some dicey problems: How can one tease apart the roles of innate capacities and learning in development? How can evidence from the behavior of nonverbal infants and animals support conclusions about the content of their perceptual experience?

The chapter ends with research by more recent computational cognitive scientists, beginning with David Marr, whose work in the 1970s complemented and expanded on the work of Gibson, Held, and others. Together, these scientists created a vibrant interdisciplinary approach to vision that bridges psychophysics, animal behavior, systems and cognitive neuroscience, artificial intelligence, and machine learning. The research this approach has fostered is beginning to suggest how innate knowledge of the visible world might arise, prior to an animal's first encounters with it. The chapter therefore introduces many of the methods and ideas that guide research on core knowledge.

Chapter 2 focuses on studies of infants' knowledge of objects: the movable bodies that we see, grasp, and act on. Before infants can reach for and manipulate objects, they organize perceptual arrays into bodies that are cohesive, bounded, solid, persisting, and movable on contact. Young infants use these abstract, interconnected properties to detect the boundaries of each object in a scene, to track objects over occlusion, and to infer their interactions with other objects. Nevertheless, there are striking limits to young infants' object representations: Infants have little ability to track hidden objects by their shapes, colors, or textures, although they do detect and remember these properties.

Above all, research reveals that infants' early-emerging representations of objects are the product of a single cognitive system that operates as an integrated whole. This system emerges early in development, it remains present and functional in children and adults, and it guides infants' learning. The system combines some, but not all, of the properties of mature perceptual systems and belief systems, and it therefore appears to occupy a middle ground between our immediate perceptual experiences on the one hand and our explicit reasoning on

the other. Research probing infants' expectations about objects suggests hypotheses concerning the mechanisms by which a system of knowledge might emerge, function, and guide infants' learning about the kinds of objects their environment provides and the kinds of events that occur when different objects interact. Research described in this chapter also reveals that infants' knowledge of objects is at least partly innate. It suggests how innate knowledge of objects might arise prior to birth, preparing infants for their first perceptual encounters with movable, solid, inanimate bodies.

Chapter 3 turns to core knowledge of places. The core place system underlies our sense of where we are, where other things are, and what paths will take us from one place to another. Studies of animals and young children reveal that navigation depends, first and foremost, on representations of abstract geometric properties of the ground surface over which we travel: the distances and directions of its boundaries, ridges, cliffs, and crevices. This research also reveals sharp limits to the features of the environment that guide children's and animals' sense of place, and it provides evidence that the place system, like the object system, is unitary, emerges early, functions throughout life, and supports learning about the navigable environment.

More than seven decades of research on place representations demonstrates how scientists can determine whether the same cognitive system exists and functions in the same manner in different animal species. The achievement is important, because studies of nonhuman animals provide a panoply of tools for probing the nature, evolution, and development of the cognitive capacities we share with them. Research provides evidence that our place system is ancient: It is largely shared by animals as distantly related to us as fish, from whom we diverged some 500 million years ago. Building on these findings, research on navigating animals, using methods of controlled rearing, provides the clearest evidence that core knowledge of places is innate.

Meanwhile, research on navigating human adults sheds light on the minds of animals and human infants. This research reveals that the place system is modular: It operates automatically, regardless of our intentions and beliefs, and its internal workings are deeply inaccessible to our conscious minds. Studies of adult humans and animals reveal that both the place and object systems are resource-limited and compete for attention. As a consequence, they are not readily combined into a unitary representation, but they complement each other. The same processes that single out movable bodies for the object system serve to remove these bodies from the representations that guide our navigation, allowing us to track our location and travels relative to the enduring and unchanging terrain over which we move. Although children and animals learn to navigate by landmarks, they do not form a comprehensive Euclidean map of their environment. Finally, research applying methods of cognitive neuroscience to studies

of animals and human adults, but not yet to studies of young children, provides evidence that navigation is aided by processes of mental simulation of different paths through the environment. Throughout this book, I speculate that core systems give rise to simulations that prepare infants for learning from their encounters with objects, places, and other entities.

Chapter 4 focuses on core knowledge of number. Research on human infants, children, adults in diverse cultures, and nonhuman animals all converges on evidence for an early-emerging ability to represent and combine numerical magnitudes with approximate, ratio-limited precision. This ability depends on a core system with most of the properties of the core object and place systems: it is present in newborn infants and functions throughout life, and it is ancient, unitary, and limited in the types of information it provides. The core number system also is modular, unconscious, and yet dependent on attention, and it competes for attention with the core system that represents the individual objects that we enumerate. Despite its modularity, studies of older children and adults provide evidence that the core number system contributes to children's learning of mathematics in school, to adults' reasoning about mathematics in everyday life, and to the thinking of professional mathematicians when they are challenged with difficult questions in diverse mathematical fields. Thus, even the most abstract and abstruse feats of human reasoning draw, in part, on ancient, early-emerging cognitive systems.

Chapter 4 completes my exposition of the general properties of core knowledge. In chapter 5, I argue that these properties go together: A cognitive system that has some of them is likely to have all of them. The most important of these properties are their ancient origins, their persistence over vast stretches of time, and their resilience despite the innumerable changes wrought by subsequent cognitive evolution. Such systems can only focus on the most abstract and highly general properties of the entities in their domain, and they can only survive within highly encapsulated brain systems, for they must function in creatures who live in very different environments, with very different brains and cognitive capacities. As the beneficiaries of hundreds of millions of years of cognitive evolution, these systems will operate with high efficiency in diverse environments.

In the next three chapters, I propose three more core cognitive systems with these properties. Chapter 6 draws on a large body of research investigating the abilities of animals, infants, children, and adults to categorize objects on the basis of their forms and functions. After reviewing a wealth of research on form perception and shape-based object recognition, I hypothesize that humans and animals are endowed with an ancient core system for perceiving, tracking, categorizing, and reasoning about the branching forms and varied functions of natural objects: especially plants. Plants are the primary sources of food for humans and other animals, and in natural environments, they remain prominent sources

of the poisons, irritants, and thorns that threaten animals' health. Like the place system, the form system centers on abstract, interconnected geometric properties, but its properties are distinct from, and complementary to, those captured by the place system.

If the form system evolved to support reasoning and learning about the shapes and functions of living kinds, however, it must be repurposed for most of the world's people, who now live in industrialized environments. This chapter therefore considers how a core form system that evolved for learning the forms and functions of plants might be harnessed by young children to support their learning about the forms and functions of artifacts. Finally, I ask whether the place and form systems together come to support children's mastery of spatial symbols. Research provides evidence that they do, but with conspicuous limits that children only overcome when they learn formal geometry in school.

Chapter 7 focuses on core knowledge of agents: beings who cause their own motion and, by moving, cause changes in the state of the world. I review research providing evidence that young human infants represent the movements of other animals and people in accord with the interconnected, abstract concepts of *cause*, *intention*, *action cost*, and *goal value*. Moreover, infants use agent representations both to guide their own actions on objects and to interpret the object-directed actions of others. Research on young infants and newborn or controlled-reared animals provides evidence that the agent system is unitary, ancient, and at least partly innate; research on older children and adults suggests that it functions throughout life, is modular, and serves as a foundation for the development of children's action planning and causal reasoning. Finally, agent representations show signature limits that distinguish them from representations of places, objects, and plants. Core representations in these four domains compete for attention as they capture complementary aspects of the environment.

Chapter 8 focuses on core knowledge of social beings: entities who endow one another with experiences like their own and who share their experiences in states of engagement. I review research providing evidence that young human infants and nonhuman primates represent both the social interactions in which they participate and those they observe, in accord with the interconnected, abstract concepts of *shareable experience* and *engagement*. Research suggests that these representations are unitary, ancient, developmentally invariant, innate, modular, and foundational for later social and moral reasoning. Adults and children view themselves and others as simultaneously social (because we engage with one another) and agentive (because we act for our own and our partners' benefit). Until 10 months of age, however, I suggest that the core agent and social systems are not readily combined. Like the core number and object systems, these systems compete for attention, allowing infants to represent people either as agents who act and cause changes in the world or as social beings who engage and share

experiences, but not both at once. Young infants lack our concepts of people as social agents, whose behavior is both social and causal, guided by mental states that are both phenomenal and intentional. I discuss the development of new concepts of people and their mental states in chapter 10.

As infants engage with other people, they begin to learn the language or languages by which people communicate with the infant and with one another. In chapter 9, I consider infants' language learning in the first year. Research reveals that infants begin to learn the rhythms and sounds of their native language even before birth. By 4 months, infants have begun to learn some of the words and phrases with which speakers convey meaning. As their learning proceeds, infants come to discern how words and part-words combine in phrases, and how speakers use these combinations to share their experiences with others. Remarkably, infants detect and use the ordering of abstract categories of words both to learn individual word meanings and to discover the specific sound contrasts that distinguish one word from another in their native language.

Research suggests that infants' language learning depends, in part, on core knowledge of the people who speak to them and core knowledge of the things and events that people talk about. The connections between language and core knowledge are especially clear in studies of newly emerging languages and of *homesign*: the gestural communication system invented by deaf children with no access to a conventional signed or spoken language. Further connections between language and core knowledge are revealed by studies of universal patterns in the functional vocabulary of mature languages. The most frequent, short, and unstressed function words and part-words in the world's languages tend to convey meanings that map to core knowledge, because adult speakers and listeners access core concepts frequently, rapidly, and automatically.

In chapter 10, I turn to two new systems of concepts that emerge, I believe, at the end of the first year. At about 10 months, infants come to combine representations from the core agent and social systems into a unitary set of concepts of people as social agents, whose object-directed actions fulfill social goals. This new conception of human action appears to arise when infants decipher the first sentences with which their social partners invite them to share attention to objects.

About 2 months later, infants come to combine representations of the action plans of agents and the shareable experiences of social beings into a unitary set of concepts of mental states as both phenomenal and intentional: states that convey people's shareable experiences of things and events. This new conception arises, I suggest, as infants decipher the distinctive meanings of the diverse content words that convey distinct perspectives on the same individuals: words like *animal, dog,* and *Rover*, applied to the same pet, or words like *give* and *take,* applied to the same exchange. These conceptual developments give infants new

ways of learning about the world and thinking about other people. They provide foundations for the prodigious learning capacities of older children.

Coda

This book is aimed at a broad community of readers. I assume no knowledge of any particular discipline, but I try to back each factual claim with evidence. I take a broad view of the evidence that bears on these claims: In each chapter, I consider how the findings of behavioral studies on infants mesh, or fail to mesh, with one another, and with experiments that test for the same capacities in other animals, in children and adults in diverse cultures, in the brains of all these creatures, and even, in some cases, in intelligent machines. To achieve this aim, I have written a longer book than you may want to read. Despite its length, however, I have had to leave out much of the beautiful work on infants that has emerged from the many fields of developmental cognitive science. I hope my colleagues will forgive me.

This book has many flaws. First, it is full of language that may mislead. Unfortunately, natural languages have no words for any of the representations that core knowledge systems deliver. When we talk to each other, what we say is informed by core knowledge, but nothing we say expresses that knowledge directly. Core knowledge is taken for granted: Its concepts and assumptions can be left unsaid. This fact has presented me with a lifelong problem: How do I talk or write about it?

The problem was first pointed out to me in 1978 by Henry Gleitman, who was then my senior colleague, one of many brilliant and valued mentors at the University of Pennsylvania, and the greatest teacher and expositor of psychology I have ever known. When I excitedly described to him the findings (by a wonderful student, Phil Kellman) of our first studies of infants' perception of objects, Henry gently but decisively deflated my claims. "Those are interesting findings, but your babies do not see objects. At best, they see schmobjects." Henry was right: Infants start out knowing almost nothing about the chairs, cups, and cars that pop into our minds when we think about objects. Similarly, infants know almost nothing about places, forms, animals, people, or number.

On every page of this book, I bend the English language in ways that invite misreadings, for I have found no happy solution to this problem. Some authors, following Piaget, adopt a new, invented vocabulary for describing the contents of infants' minds, but as a student, I found his writings about "schemas," "assimilation," and "secondary circular reactions" less illuminating than his beautifully clear descriptions, in plain language, of his infants' responses to the challenges their investigator-parents presented to them. Here I opt for ordinary language,

but you are welcome, reader, to think "schmobject" or "schnumber" as you proceed through this book.

Second, each chapter of the book is dense with information, twists, and turns. Here is why: I aim to describe what we learn when we study the minds of infants. The first thing we learn, in embarking on such studies, is that our intuitions about infants' minds are wrong. To learn how infants think, we have to listen to what the infants in our studies tell us. This book is a portrait of the lessons learned from infants who have responded to the often-misguided questions that psychologists like me have put to them. Within this body of research, there are no "silver bullet," stand-alone experiments that capture the contents of infants' minds. Insights come instead when a chorus of voices, from the participants in many studies, starts to harmonize, revealing how infants are construing the events experimenters present to them.

Can one digestible book describe the minds of infants for a critical and curious reader with finite time, who will want to know what infants do in laboratory experiments, and to consider whether their behavior in these experiments supports the conclusions I draw? I hope this book achieves that goal, but in case it doesn't, each of its chapters—exploring infants' initial and developing knowledge of objects, places, number, forms, agents, social beings, language, and social agents—begins with a road map of the research to be presented and ends with a portrait of the cognitive capacities the research illuminates. The heart of the work lies in between, where specific experiments address specific questions that together converge on a more general understanding of infants' minds. You, reader, may choose how deeply to delve into this material.

Although this book and its successor address questions that are straightforwardly empirical—Where does knowledge begin in human infancy? How does it grow in childhood?—a passion and a hope lie behind the effort to answer these questions. By gaining a better understanding of infants and children, we gain insight into our own minds: how all people are the same and how we differ; what aspects of our thoughts and actions can be changed and what aspects we likely must live with forever; and how we can adapt our actions to leverage our cognitive capacities effectively. By coming to understand our own minds better, we will better equip ourselves to deal with the prodigious set of challenges that our endlessly inventive species has created.

Elizabeth Spelke
Cambridge, Massachusetts

1
Vision

When we look around, we see a stable and extended layout of surfaces, furnished with objects, animals, and people. Because vision brings us high-quality information about things at a distance, it is the primary sense by which we guide our actions: We don't need to wait until an attacking predator reaches us in order to decide to run for cover. Yet the capacity to see distant objects and surfaces puzzled philosophers and scientists for millennia. All perception depends on events that take place at the receptors in our eyes, ears, and other sense organs. How do such local and limited sensory events give rise to experiences of rich and varied three-dimensional (3D) scenes?

This question often leads to another: Do any of our visual perceptions depend on mechanisms that grow within us, in accord with the structure and timing of their own intrinsic processes, or do all visual mechanisms come into being through the shaping effects of our encounters with visible 3D environments? Debate over this question, dividing nativist from empiricist theories of perception, resounded through centuries of scientific and philosophical inquiry, but a large piece of the question, focused on perception of surfaces and things at a distance, was settled in the second half of the twentieth century. The research that ended this corner of the nativist-empiricist controversy sheds light on the nature, origins, and evolution of our perceptual systems, on the nature and course of perceptual learning, and even on the core functions of vision. It also provides us with tools for investigating the nature and origins of our more properly cognitive systems, including the systems by which we keep track of an object that has moved completely out of view, or reason about the imperceptible intentions behind a person's actions.

At the center of these achievements stands the work of two twentieth-century psychologists, Eleanor Gibson and Richard Held. My first chapter describes their discoveries at some length, because their work bears on three questions that will recur throughout this book. First, what do claims of innateness and learning mean and how can research address them? Second, what cognitive capacities are common to people of all ages, and what capacities emerge or change fundamentally over development? Third, what cognitive capacities are unique to humans, and what capacities are shared by other animals? The empirical and conceptual tools that were developed to address questions of innateness, of continuity over ontogeny, and of continuity over phylogeny will provide foundations for all the

What Babies Know. Elizabeth S. Spelke, Oxford University Press. © Elizabeth S. Spelke 2022.
DOI: 10.1093/oso/9780190618247.003.0001

research covered in this book, probing the origins and early growth of knowledge in domains from physics to mathematics to morality.

I bracket the tale of these two scientists with briefer tours through three other moments in our intellectual history. I begin with a seventeenth-century controversy between two philosophers—René Descartes and George Berkeley—over the nature of vision (section 1). Then I turn to the psychophysical methods, developed in the nineteenth century, that transformed the study of vision from philosophy to science. I focus on the work and thinking of the physicist, physiologist, and vision scientist Hermann von Helmholtz, who combined Descartes's and Berkeley's most fruitful empirical claims (section 2). After discussing the work of Gibson and Held (section 3), I end with a sweep through selected research in computer science and neuroscience (section 4). I focus on David Marr's key insight, building on his efforts to get computers to solve a variety of visual tasks solved by humans and animals, that the central function of vision is to deliver, to more central cognitive systems, a representation of the continuous surface layout that reflects light to the eyes. I also build on insights from generations of neuroscientists and computational cognitive scientists, and consider how patterns of simultaneous, intrinsically generated activity in networks of neurons may prepare perceivers for the environments and activities that they come to experience. And I end with brand-new research, suggesting that these insights may combine to lead future developmental cognitive scientists to a deeper understanding of the precocious perceptual abilities revealed by Gibson, Held, and their descendants.

I focus in this chapter on the origins and development of perception of the visible surface layout, leaving other topics that traditionally are seen as part of vision—including our abilities to represent the forms and functions of objects and the faces and actions of other people—for later chapters. Research on visual space perception sets up the main project of this book, in three ways. First, it illustrates how claims of innateness and learning can be turned into scientifically tractable questions. Investigators can study the perceptions of human infants and nonhuman animals much as we study our own perceptions. Although no person will ever have direct access to the mind of another individual of any age or species, Helmholtz's psychophysical methods give us indirect access to the minds of our peers, and Gibson's and Held's variations on those methods give us similar access to the minds of infants and nonhuman animals. Thus, studies of newborn animals, of animals raised under controlled conditions, and of the developing neural systems that support perception can probe systematically the effects, and noneffects, of experience on perceptual development.

Second, the study of visual space perception has been profoundly and thoroughly interdisciplinary for centuries. Descartes's essay on vision synthesized

studies of the physics of light, the biology of the eye, the psychology of perception, and the mathematics of Euclidean geometry. Helmholtz's great treatise on vision referred to the science of vision as "physiological optics," merging psychology, biology, and physics. Marr argued explicitly that an understanding of vision requires an interdisciplinary synthesis of neurobiology, psychophysics, and the mathematics of logic and computation. Vision science has grown as these contributing disciplines have advanced and interconnected. In future chapters, I will focus first and foremost on studies using behavioral methods, like those of Helmholtz, Gibson, and Held, to investigate cognitive functions beyond perception, and I will interpret the findings of these experiments in the context of the same interdisciplinary synthesis exhibited in this chapter. Interconnected studies of minds, brains, and machines bring insights into all areas of developmental cognitive science.

Third, the study of vision shines a light on the territory that the rest of this book will cover, by revealing the limits to what we can know purely by seeing, and by characterizing some of the most important representations that core knowledge systems build on. The study of vision and its limits helps us to understand the functioning of the core cognitive systems discussed in this book, whose tasks begin where visual perception ends.

1 Descartes, Berkeley, and the Philosophy and Science of Spatial Vision

In *Optics* (1637/2001), Descartes proposed that humans and animals see distance and direction "as it were, by a natural geometry," instantiated in mechanisms that infer the position of each region within a visual scene from geometrical relationships between the eyes and light-reflecting surfaces. When a perceiver looks at a point on a visible surface, for example, the point's location can be computed from the distance separating the two eyes and the direction of each eye, just as the position of the third corner of a triangle follows from the angular sizes of the other corners and the distance between them (figure 1.1a). The natural geometry of the visual system gives rise to the depth cue of *convergence*: The more the eyes turn inward, the closer the point to which they are directed (figure 1.1b). Descartes suggested that this relationship is not specific to visual experience: A blind man holding two sticks might, by the same rational inference, perceive locations at a distance by touch (figure 1.1c). He showed how abstract, geometric computations account for a number of other visual depth cues, for the perception of the constant sizes of objects over changes in their distance, and for illusions of size perception, like the perceived changes in the size of the rising moon, whose distance we register as greatest at the horizon: an illusion that our

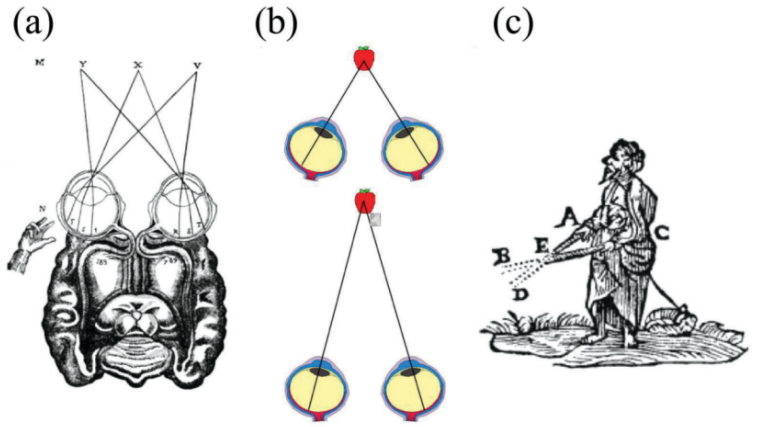

Figure 1.1 The depth cue of convergence, illustrated by analogy with touch. When a perceiver looks at one object with both eyes, the object's distance can be computed from the distance separating the two eyes and the angular direction of each eye (a and b). This process resembles a process for inferring the distance at which two sticks cross, without vision, as one holds each stick in one hand (c). (Images a and c are reproduced from Descartes, 1637/2001.)

mechanisms of depth perception, which function to detect distances on Earth, are powerless to correct.

Descartes's account of vision has three key features. First, it is mechanistic: Our visual experience depends on our sensory organs, on the geometric properties of light, and on the neural mechanisms that respond to this stimulation. Second, his account does not appeal to learning; instead he suggests that the extraction of geometric properties from visual stimulation is "natural" to humans and animals. Third, he proposes that the process giving rise to depth perception has the form of a rational inference. The claim that humans generate new knowledge by combining representations of abstract properties in accord with logical rules, independently of any encounters with external things, is paradigmatic of rationalist, nativist theories in cognitive science.

In his *Essay Towards a New Theory of Vision* (1709/1975), the philosopher George Berkeley denied these claims and offered a different view. He focused not on the mechanisms of visual perception but on our conscious visual experiences. When we see a spatial layout, he reasoned, we don't experience lines and angles or work through a sequence of steps in a logical syllogism: We experience the whiteness of the couch and the redness of the pillow that sits on it. Perceptual experience begins not with abstract geometry but with modality-specific sensations, and it grows not by rational inferences but by associative learning. When we

experience distinct sensations at the same time or in immediate succession, we come to link them. In the case of convergence, viewing an approaching object with both eyes elicits concurrent visual sensations and sensations from the muscles that move the eyes. As the object gets closer, these visual and muscular sensations become more intense as the eyes converge further and further inward, away from their natural state of rest. As infants, we learn to associate these visual and muscular sensations with one another, and also with tactile sensations that occur on contact with the object. By associating diverse sensations within and across modalities, children come to experience a unitary, 3D environment.

Much of the disagreement between Descartes and Berkeley stemmed from their differing philosophical claims concerning the nature of the world and of human knowledge. Those differences may have led them to different views of the nature of perception: as an immediate, conscious, sensory experience for Berkeley, and as a mechanistic process that captures the geometry of the visible environment in a quasi-rational manner for Descartes.[1] Nevertheless, their argument over the origins of depth perception appears to be straightforwardly empirical. According to Descartes, visual perception of depth should occur the first time that an infant or animal with a functioning visual system encounters a visible environment, because neither geometric representations nor rational inferences are learned.[2] According to Berkeley, in contrast, visual perception of depth will only emerge after repeated experiences in a visible environment in which the critical associations can form.

In letters written to Locke, Molyneux (1688/1978; 1693/1978) proposed experiments that aimed to resolve this controversy. His first letter asked, if a man born blind and able to reach for objects by touch were to have his blindness corrected, "Could he know by his sight, before he stretch'd out his hand, whether he could not reach them, tho they were removed 20 or 1000 feet from him?" Locke did not respond to this letter (Degenaar & Lokhorst, 2011), but Molyneux (1693/1978) later raised a similar question concerning such a man's perception of object shape, and Locke responded with enthusiasm in his *Essay*

[1] Descartes viewed vision as a process that is common to humans and animals, but he believed that only humans are endowed with the capacity for true reason. Contra Berkeley's caricature, Descartes could not have meant that infants, animals, or adults consciously engage in geometric reasoning when they see depth. Nevertheless, Descartes did not elaborate on the sorts of geometric representations and logical operations that visual space perception requires.

[2] For convergence to specify the absolute distance of an object, one must know the distance between the two eyes of the perceiver: a quantity that varies with growth. In the absence of this knowledge, however, a newborn perceiver could use convergence to perceive the relative distances of different objects: When one looks with both eyes at objects of different distances, the directions to which the eyes are pointed change, but the distance between the two eyes does not (figure 1.1b); if the eyes turn inward farther for one object than for the other, therefore, the first object is closer to them. If absolute distance is measured in bodily units such as the length of the arm, moreover, convergence will provide an approximate specification of absolute distance in body coordinates, as the eyes grow farther apart in rough proportion to the growth of the limbs (von Hofsten, 1982).

Concerning Human Understanding (1690/1975). Neither of Molyneux's proposed experiments was possible to perform at the time, however, so the discussion remained within philosophy.[3] I will jump, therefore, to the nineteenth century, when debates over the origins of visual space perception moved from philosophy to science.

2 Helmholtz and the Emergence of Psychophysics

The nineteenth century was a fruitful time for the cognitive and brain sciences in general, and for the nativist-empiricist controversy in particular, but I focus on the work of just one scientist. Hermann von Helmholtz was a physicist, neuroscientist, and brilliant experimental psychologist who probed unseen neural processes through behavioral experiments. As a student of physiology, he devised a method for measuring the speed at which signals travel along nerves (many decades before more direct measurements of neuronal activity could be made) by systematically comparing the latencies at which a frog's leg muscle contracted after the nerve was stimulated at different distances from the spinal cord.[4] Later, Helmholtz showed similar genius as an investigator of central aspects of auditory perception of tone and visual perception of color. Here, however, I consider only his work on visual space perception.

By the mid-nineteenth century, a new experimental method had emerged that combined Berkeley's focus on conscious visual experience with Descartes's focus on visual mechanisms. Experiments in *psychophysics* probe the mechanisms that give rise to perception by exploring the relations between systematic changes in the optical stimulation that an observer receives to changes in the perceptions that he experiences. Using this method, additional mechanisms of visual depth perception had been discovered and studied systematically by vision scientists, using themselves as the observers. The findings of these studies were synthesized in Helmholtz's landmark treatise, the *Treatise on Physiological Optics*

[3] Interestingly, one of Held's last articles addressed Molyneux's second question (Held et al., 2011). After surgery restored the sight of adolescents who were blind from early in life, the youths were presented with unfamiliar objects which, like factory-assembled artifacts, were composed of blocks and cylinders in shapes like those formed by Legoset. The youths discriminated between two objects with different arrangements of parts by touch alone and, to some degree, by vision alone, but shape perception in one modality did not support recognition in the other modality. Because the experiment focused on shape properties that became relevant to humans only after the industrial revolution, and because children become sensitive to these properties only after infancy (see chapter 6), I discuss this research in *How Children Learn*.

[4] This study was an important precursor to research in cognitive psychology, conducted a century later, that used similar chronometric methods to measure the processes by which people attend to, compare, categorize, or mentally rotate visible objects (Neisser, 1967; Shepard & Metzler, 1971; Posner, 1978).

(1867), combining ideas from the Cartesian and Berkeleyan traditions. Like Berkeley, Helmholtz rooted spatial vision in depthless sensations; like Descartes, he argued that mature visual perception of the surrounding layout depends on geometric information and takes the form of a rational but unconscious inference. Helmholtz conducted psychophysical experiments on himself as he looked at and acted on visual displays, to study diverse aspects of visual space perception, including the perception of stereoscopic depth, which I discuss later in this chapter, and the perception of visual direction. I focus here on the latter ability.

In the last volume of *Physiological Optics* (1867, Vol. 3) Helmholtz used psychophysical methods to ask whether his own visual perception of distance and direction depended on mechanisms that are fixed or are more flexible. He tested the effects of wearing prism lenses that displaced the light reflected to the eyes, initially causing objects to appear to the left or right of their actual positions. On first looking through the prisms, Helmholtz verified that when he looked at an object and then closed his eyes and pointed to its seen position, his pointing was displaced (figure 1.2a). After repeatedly pointing at objects and noting these errors, however, the errors diminished (figure 1.2c). Moreover, if he removed the prism lenses, attended to the object, and then pointed again without vision, he observed an opposite pointing error. Thus, the change was not a conscious, strategic adjustment to wearing prisms (figure 1.2b).

These experiments showed that at least some aspect of space perception is malleable in adults, but they do not address the origins of space perception: Do we experience depth on our first encounters with a visible layout, or do we learn to relate our varying visual, tactile, and muscular sensations to one another? Helmholtz doubted that this question could be answered by psychophysical methods, which rely on carefully trained observers. "Little enough is definitely known about infants and very young animals, and the interpretation of such observations as have been made on them is extremely doubtful" (Helmholtz, 1867, vol. 3, ch. 26, p. 5). He suggested that empiricist theories were more likely to be correct on grounds of parsimony, however, for two reasons. First, his experiments showed that space perception is indeed modifiable in adults, in response to changes in visual experience. Second, because the eyes grow progressively farther apart from infancy to adulthood, an innate specification of depth from convergence, which requires a commitment to a particular interocular distance, might be a hindrance, and the learning capacity found in adults likely would be especially useful during development. Thus, he concluded, the mechanisms by which adults adjust to prisms likely account for the development of visual space perception as well.

Nevertheless, arguments from parsimony rarely settle questions in cognitive science, and Helmholtz was cautious in the conclusions he drew from them. In retrospect, his caution was prescient, because subsequent research has

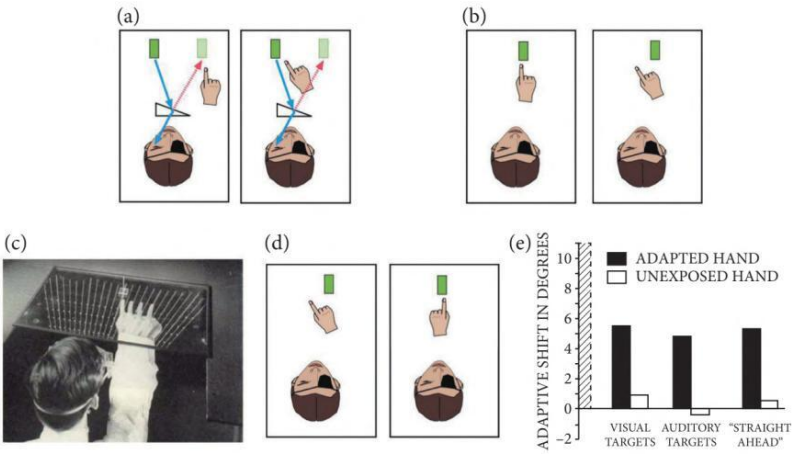

Figure 1.2 Prism adaptation experiments to test the malleability of visual perception of direction. Immediately after covering one eye and placing a prism in front of the other eye, attempts to point to a visible object, without sight of the hand, lead to errors (a, left). After experience pointing while wearing the prism, however, the error diminishes (a, right, and c). This change does not reflect a conscious judgment, because when the prisms are removed, the opposite error occurs (b, right). This after-effect became the standard measure of prism adaptation (e.g., e, dark bars). In research by Harris (1963), however, the after-effect occurred only when the perceiver pointed with the hand used during adaptation (d, left), not the other hand (d, right, and e, left bars). The same findings were obtained when perceivers pointed to auditory targets (e, middle bars) or straight ahead to no target (e, right bars). Thus, adaptation does not produce a change in visual perception. (Images c and e are reproduced from Harris (1963). I thank Kirsten Condry for images a, b, and d.

overturned two assumptions underlying his interpretation of prism experiments. First, Helmholtz posited that the experience of pointing to objects while looking through a prism produces a change in *visual* perception: a change in the seen position of the object, rather than in the felt position of the hand with which he pointed. He based this assumption on the findings of further experiments in which he pointed to objects seen through a prism with just one arm until his pointing errors declined, and then tested his pointing with the other arm. He reported that experience reaching with one arm led to a change in reaching with the other arm, suggesting that the adaptation was caused by a change in vision rather than in proprioception.

Helmholtz did not control for the movements of other body parts, however, and later experiments that did so yielded different findings (Harris, 1963, 1965;

figure 1.2c). When adults point to a seen object while moving one arm but not the head or body, adaptation to wedge prisms produces an aftereffect in the arm that did the pointing but not in the other arm (figure 1.2d). In addition, adults who have adapted to prisms show a change in pointing not only when they point to visible targets but also when they point to unseen sound sources, when they point straight ahead, or when they use one hand to point to the other (figure 1.2e). Adaptation to prisms therefore produces a change in where perceivers feel their hands to be, not a change in the perceived positions of visible objects. Although space perception indeed is modifiable in response to changes in visual stimulation, it is proprioception, not vision, that is modified.

The second assumption, made by Helmholtz and overturned by later research, concerned the mechanisms by which experience alters the relation between spatial vision and proprioception. Centuries of empiricists proposed that relations between vision and touch were learned in a piecemeal fashion, in which each visual position is separately linked to a corresponding position in motor space. On such a view, learning is likely to be slow (because each new link between a seen and a felt position is learned separately) but flexible (because any pattern of linkages can be learned, as long as it presents a consistent mapping between seen and felt positions).

Research now reveals, however, that relations between vision and touch are more constrained, and learning is fast. When perceivers view a target object at just two locations, one on the left and the other on the right, they show adaptation not only at these locations but at new locations, in directions that were never visible during prism viewing (Bedford, 1989; Ghahramani, Wolpert, & Jordan, 1996; Ghahramani & Wolpert, 1997; Faisal & Wolpert, 2009; figure 1.3a). If the two points move leftward or rightward, then perceivers show leftward or rightward changes of roughly equal magnitude in pointing at all directions within the space containing those locations; if the two points move outward or inward, then perceivers' pointing reveals a roughly uniform expansion or contraction of the space between those points, though not outside them (figure 1.3b).

These findings indicate that observers don't just learn about the locations they have seen and touched—their learning generalizes to other locations that are consistent with a more orderly change in the mapping between vision and touch. Further work shows, moreover, that observers cannot adjust to all possible changes. After training at points whose displacements are inconsistent with any uniform displacement, expansion, or contraction, observers fail to learn the transformation, even at the points that they have been trained on. Instead, they learn an approximation to the best-fitting global change of direction and scale that applies to all locations between those points (figure 1.3c). Adults adjust to prisms by applying relatively global transformations to the function that generated their previous mapping between vision and touch.

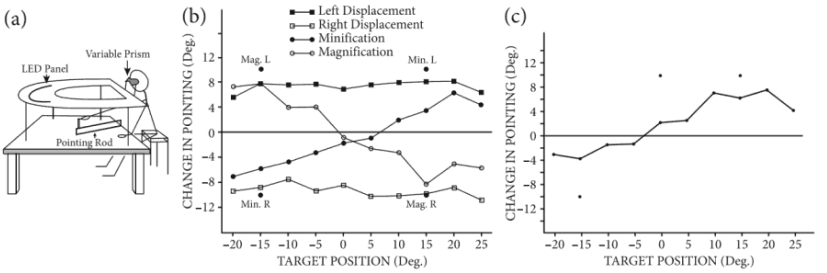

Figure 1.3 A prism adaptation experiment in which perceivers view targets that appear at just two or three locations, displaced from their true positions. In (a), perceivers can move freely, but the target is visible only at the specified positions. In (b) and (c), the position and displacement of each target at test, after removal of the prisms, is indicated by dots, and the change in participants' pointing both to trained and to untrained positions is indicated by lines. When targets appear at two locations, adaptation generalizes to locations between them, resulting in a uniform displacement or scaling of that region of the visual field (b). When three visible target locations are not consistent with any uniform transformation (c), perceivers extrapolated a transformation that combined a leftward displacement with a minification of the visual field. Note that the transformation is not strictly linear. (Reproduced from Bedford, 1989.)

This research suggests that the capacity to adjust to changes in the mapping between vision and touch itself requires a preexisting mapping between these modalities: it does not support learning of any arbitrary mapping de novo. If one believes, with Helmholtz, that a theory positing a single mechanism of perceptual learning, operating throughout human development, is more parsimonious than a theory positing two distinct mechanisms (one that operates in infancy and a second that operates later in life), then these findings turn his empiricist conclusion on its head. Infants would begin life with an unlearned, geometrically regular mapping between visual space and proprioceptive space, and subsequent experience would modify this mapping in a restricted set of ways. These findings do not settle the nativist-empiricist debate over visual space perception, however, because future discoveries may undermine this parsimony argument as well.[5] Indeed, Helmholtz expressed some pessimism as to whether the origins of space perception would ever be elucidated. I will spend some time unpacking

[5] In comments on an earlier version of this chapter, Tomer Ullman pointed out that current discoveries may have already done so: Multilayered convolutional neural networks, a modern instantiation of empiricist ideas that incorporates aspects of the spatiotopic organization of biological visual systems, may be biased by their architecture to learn only the kinds of relationships learned by participants in the prism adaptation experiments by Bedford and Ghahramani.

his grounds for pessimism, because they challenge all the work to be discussed in this book.

I begin by considering Helmholtz's central method for studying perception: psychophysics. Just as he studied the quantitative relationship between the position of stimulation of the frog's leg and the latency of its retraction to shed light on the underlying mechanism of nervous conduction, he studied the quantitative relationship between the physical stimulus presented to a perceiver and the perceiver's judgment or action to shed light on the underlying mechanisms of visual perception. Using psychophysical methods, Helmholtz explored in exquisite detail the mechanisms that give rise to the conscious experience of tone, color, and depth, by documenting systematically how his perceptual experiences varied with changes in the stimulus array.

Helmholtz noted, however, a critical difference between his studies of the frog's nervous system on the one hand and of his own conscious perceptions on the other. The latter studies probed relationships between physical stimulation and reportable, phenomenal experiences. Many findings from these experiments form the foundations of our understanding of perception today, but they have a critical limitation. If psychophysics depends on reports of conscious experiences, then experimenters can only draw conclusions by experimenting on themselves. Psychophysical studies can be generalized from one person to another, if different scientists perform the same experiments and report the same findings. They do not generalize, however, to newborn human infants or animals, who cannot perform psychophysical experiments or report their findings. Scientists can study what sensory events infants and animals respond to, but their findings won't tell us what infants and animals *perceive*.

Helmholtz's problem cuts deep. Consider some experiments from the 1950s on recently hatched chicks and newborn human infants. The psychologist Robert Fantz presented chicks—raised in darkness from the time of hatching until the onset of the experiment—with two trays of small objects. One tray presented 3D objects of the sizes and shapes of grains. The other tray presented two-dimensional (2D) images of those objects: flat discs. When chicks were allowed to go to either tray to peck for food, they went for the 3D objects (Fantz, 1958). Fantz then adapted the choice experiment for human infants by presenting them with a large sphere and a flat disc of the same color and circumference, side by side. Infants looked more at the sphere (Fantz, 1961). Based on these and related findings, Fantz concluded that space perception arises in the absence of visual experience, both in birds and in humans.

Fantz was an early user of two experimental methods—studies of the spatial behavior of dark-reared animals, and studies of the visual exploratory behavior of human infants—on which I will rely heavily in later chapters. Nevertheless, there is a problem with the conclusion he drew from these studies. As the

research of Helmholtz and many others had shown, perception of a 3D object arises when the visual system detects particular patterns of stimulation, such as the oculomotor adjustments that occur when the eyes converge on a point at a particular distance. For adults, these sensory patterns give rise to perception of depth, but for chicks and infants, they might simply be experienced as visual and muscular sensations, as Berkeley proposed. Do chicks and babies choose spheres over discs on the basis of their perceived 3D shapes, or on the basis of the visual and kinesthetic sensations they elicit?

This question does not reflect an abstruse academic quibble but a serious disagreement over the origins of space perception. Consider first Fantz's experiment with human infants. If infants are endowed with a Cartesian natural geometry, then participants in Fantz's experiments, who converge their eyes on different parts of a sphere, would perceive its 3D shape, and they might well prefer to look at a sphere than at a disc, because it has a richer geometric structure. Alternatively, if infants experience only a distinctive complex of visual and muscular sensations, as generations of empiricists have proposed, then the participants who look at different regions of Fantz's displays will have richer, more variable sensory experiences when they view the sphere than when they view the disc, because the variable distances of points on the sphere will lead to variable experiences of convergence. Infants therefore should prefer looking in the direction of the display that elicits these experiences, even if they fail to perceive its depth. Because the infants themselves cannot report on their experiences, we do not know which experiences they were having.

Now consider the chick. Unlike human infants, chicks are precocial animals that engage in coordinated, adaptive behavior as soon as they begin to look around and move: They peck at round grains, not at flat ones. Because eating is necessary for survival, natural selection evidently ensured that their actions are guided by a mechanism that begins to operate when the chick first sees peckable food objects. But what rules govern the operation of that mechanism? If chicks have an innate capacity for depth perception, then their pecking may be guided by a rule such as "peck the spheres" or "peck the solid grains." If instead they learn to perceive depth and begin by experiencing only visual and tactile sensations, then their early pecking may be guided by a rule such as "peck where the changes in visual and muscular sensations occur." Either mechanism would suffice to produce the adaptive behavior Fantz observed.[6]

[6] In an insightful discussion, Burge (2010) argued that some of the alternative rules philosophers have proposed to account for the visually guided behavior of nonlinguistic animals or prelinguistic infants can be rejected, because the entities to which they appeal play no role in scientific explanations of animal or human life. For example, psychologists need not worry whether Fantz's chicks construed each peckable display as a single grain or as a collection of undetached grain parts (in reply to Quine, 1960), because representations of the latter sort do not enter into any accounts of the adaptive activity either of chicks or of older members of their species, past or present. Even if Burge's argument is

With no access to the phenomenal experience of the newborn infant or newly hatched chick, we are back to the original question at the heart of the nativist-empiricist debate: Does visual perception of depth arise through the shaping effects of experience, or is it present in inexperienced infants and animals? Do infants and animals experience a mix of depthless sensations or a stable 3D layout? This question, raised in one form or another for more than two millennia, clearly is not easily answered. Nevertheless, Eleanor Gibson and Richard Held began new attacks on it in the 1950s, extending Helmholtz's psychophysical methods to perceivers who cannot tell us what they are experiencing and devising new research strategies that addressed Helmholtz's problem.

3 Testing the Origins of Space Perception

Before Gibson studied the perceptual abilities of infants and animals, she studied the psychophysics of depth perception in adults. With James Gibson, she tested military recruits who were training in natural, outdoor environments. Natural ground surfaces, such as fields of randomly scattered grass, project gradients of texture to the eyes, with texture elements that become progressively smaller and more densely packed as the ground surface recedes. As a perceiver moves through the environment, the images of these texture elements are displaced in her visual field in regular patterns of optic flow: If she moves over open terrain, her goal remains centered in the visual field, and surrounding objects and texture elements expand outward toward her visual periphery (J. J. Gibson, 1950). Moreover, the speed of the visible movement of objects that are located in the same direction depends on their distance from her, just as the images of distant trees move slowly, and of nearby trees move rapidly, when we view a wooded landscape through the window of a moving train. The Gibsons hypothesized that adults perceive the distances of locations of this terrain by detecting these patterns of optic flow. Consistent with this hypothesis, cadets who were allowed to move their heads and bodies freely as they stood on open, grassy terrain perceived the distances of landmark objects placed on that terrain quite accurately (Gibson & Bergman, 1954/1991).[7] In the laboratory, moreover, when 2D images

accepted, however, it does not decide between the two alternative rules sketched here. Visual systems represent both solid objects and patterns of double images, and both these representations figure in explanations of action and perception (in particular, the differences in the images projected to the two eyes from a 3D array figure in explanations of stereoscopic depth perception). Neither of these two rules therefore can be rejected out of hand: a choice between them must be made on the basis of evidence.

[7] Binocular information is especially useful for detecting depth differences in nearby objects; it plays little role in perception of the more distant, large-scale surface layout.

of a textured surface were projected on a single flat screen while the original surface was rotated or translated in depth, adults who observed the moving texture on the screen reported compelling perceptions of surface depth and motion (Gibson & Gibson, 1957/1991).

These experiments were conducted with human adults, but Gibson had long been interested in animal behavior and learning. In the early 1950s, she observed that newborn goats, placed on a tall but narrow stool immediately after birth, remained there calmly until an experimenter removed them to an extended surface, whereupon they began to walk (Gibson, 2002). This behavior suggested that the newborn animal used some information from its senses to guide its locomotion—but information from which senses, and of what sort? Was the inexperienced goat's walking guided by the sight, scent, or touch of the floor? And if it was visually guided, did the goat perceive depth in the same manner as the soldiers, from patterns of optic flow? Working with two new collaborators, Richard Walk and Thomas Tighe, Gibson addressed these questions by creating one of the most beautiful experimental paradigms in the history of psychology: the "visual cliff" (Walk, Gibson, & Tighe, 1957/1991; Gibson & Walk, 1960; Walk & Gibson, 1961).

The visual cliff consisted of a narrow, raised centerboard flanked by two rigid, transparent surfaces (figure 1.4). Directly beneath one of the surfaces, Gibson placed a textured pattern of the sort used in her studies of depth perception in adults: Looking down from the board on this side, we see a nearby textured surface of support. Several feet below the other surface, Gibson placed the same textured pattern on the floor: looking down on that side, we see a drop-off to a distant surface. When Gibson and her collaborators placed newborn goats on

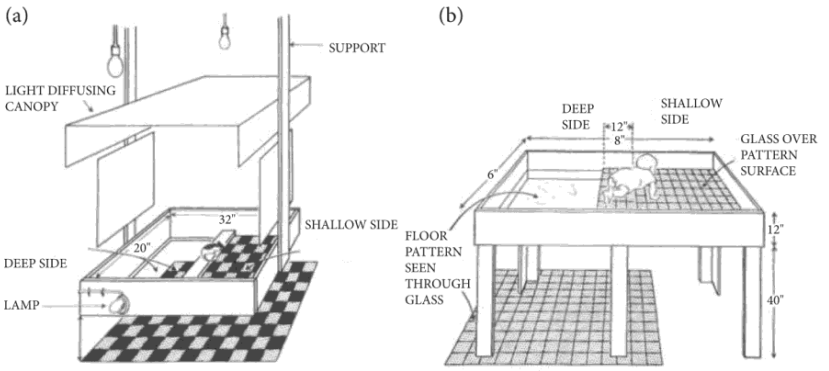

Figure 1.4 Gibson's visual cliff for (a) small animals, and (b) human infants, who avoided the deep side as soon as they began to use vision to guide their locomotion. (Reproduced from Walk & Gibson, 1961; reprinted in E. J. Gibson, 1991).

the centerboard, they walked onto the shallow side only, avoiding the visual cliff, even though the transparent surfaces on the two sides felt the same (E. J. Gibson & Walk, 1960). Her first question was answered: information from vision, not touch, keeps newly born goats from wandering off a stool.

Goats are mountain animals who locomote independently from birth, so it was possible that their avoidance of drop-offs depended on a species-specific talent. To test the generality of this ability, Gibson and her collaborators tested a variety of animals, including rats, cats, and human infants, who begin life in protected environments and do not engage in visually guided, independent locomotion until well after birth. As soon as these animals began to locomote (around 7 months of age for human infants), they used vision to guide them onto nearby extended surfaces, avoiding visible drop-offs (Walk & Gibson, 1961). Experiments on rats, using a dark-rearing method like that of Fantz (1958), showed that cliff avoidance was unaffected by visual experience: Rats reared in the dark avoided the cliff on first exposure to a lighted environment (Walk et al., 1957/1991). This finding doesn't tell us, however, whether rats and goats perceive the 3D spatial layout as humans do.

Accordingly, Gibson next asked what kind of visual information guides the locomotion of newborn goats and dark-reared rats. Across a series of experiments manipulating the visual information for depth, Walk and Gibson found that all the animals they tested avoided the deep side of the cliff by using the gradients of motion, or optic flow, that give rise to visual space perception in human adults (Walk et al., 1957/1991; Walk & Gibson, 1961; Walk, 1968; for a review, see Gibson, 1969). Avoidance of the deep side of the cliff was equally robust in animals who viewed the array with one versus two eyes, eliminating binocular information for depth. Moreover, cliff avoidance was not affected by the so-called pictorial cues to depth, used since the Renaissance to lend an impression of depth to 2D images: when the two sides of the cliff showed flat images at the same distance but presenting forms of different relative sizes and spacing, animals crossed them equally.

Thus, the optic flow used by animals and human infants on the visual cliff is the same as that which evokes perception of an extended 3D, outdoor surface layout in mature, articulate perceivers. With these findings, Gibson created a bridge between psychophysical studies of human adults, who can report on their conscious perceptions of surfaces that vary in distance, and studies of nonverbal animals, who traverse or avoid such surfaces. The common pattern of findings across these experiments provides the first suggestion of common mechanisms for visual perception of surface layouts across species and ages. Indeed, the convergence Gibson found between the psychophysical functions relating stimulus information to spatial behavior in goats, rats, and human adults looks a lot like the convergence Helmholtz called for between the psychophysical functions

relating physical stimulation to reports of sensory experiences by different human experimenters.

Further studies by Gibson revealed an interesting difference between goats and rats on the one hand and cats on the other (Gibson & Walk, 1960). When kittens were reared entirely in darkness and then placed on the lighted visual cliff, they moved around at random during their first day in the light, crossing equally onto the deep and shallow sides of the cliff. Visual experience influenced, in some way, the behavior of cats on the visual cliff.

What is the nature of this influence? Walk and Gibson (1961) considered two possibilities. First, kittens may learn, by moving and looking around, that certain patterns of visual information specify a drop-off: They may look over an edge, start to walk forward, and thereby discover, as they lose their footing, that the surface is far away. Second, perception of a drop-off may be unlearned, but experience may influence the use of this mechanism in guiding the cats' behavior. During the month of dark rearing, for example, the kittens may have learned to rely on their whiskers to guide their locomotion, and this reliance may have continued during their first day on the lighted cliff. On the second account, dark rearing would influence the *expression* of the cat's capacity for depth perception, but not the development of that capacity.

To distinguish these possibilities, Gibson and Walk (1960; see also Gibson, 1991) restricted the visual experience of their dark-reared kittens *to the visual cliff itself*: The time they spent on the visual cliff apparatus provided their only opportunity for locomotion across visible surfaces. Because the kittens were comfortably supported by the same transparent surface throughout their exploration on both sides of the cliff, all their experiences indicated that both sides of the apparatus were safe. If cats learn to avoid visual cliffs by experiencing the relationship between seeing a precipice and losing support, then these kittens should have learned to walk happily off the centerboard in both directions, because neither direction resulted in a fall.

The results showed otherwise. Despite their experience locomoting safely on the transparent glass on both sides of the centerboard, the dark-reared kittens started to avoid the visual cliff once they became accustomed to a lighted environment, just as surely and consistently as the light-reared animals did. Walk and Gibson concluded that visual experience does not serve specifically to teach cats about the dangers of visible drop-offs, by any reasonable notion of learning.[8] Despite large differences across species in the timing of the emergence of visually

[8] Walk and Gibson (1961, p. 39) write: "The animals had equal experience descending on the two sides in the beginning; according to a reinforcement learning theory, they should have learned that descent to *either* side was perfectly safe—the glass surfaces were identical tactually and kinesthetically. How is the visual difference learned, if not by confirmation? And if one supposes it is learned

guided locomotion, Gibson's controlled rearing experiments uncovered a profound commonality in the visually guided actions of different animal species.

For all their elegance, however, these experiments left two questions unanswered. First, what *are* the effects of experience on cliff avoidance in the cat: Why did dark rearing suppress the cats' use of vision to guide their first steps in the light? Second, what do these studies of cats say about the development of space perception and spatially appropriate actions in humans? Human infants begin to locomote independently at much later ages than do the other animals Gibson tested, and they cannot ethically be studied by controlled rearing methods. Shortly after Gibson began her visual cliff experiments, Richard Held and his student and collaborator Alan Hein began to conduct the research that would speak to these questions.

Held and Hein's experiments began with the finding, described earlier, that adults wearing prisms can quickly learn to adjust to them and reach accurately for visually displaced objects. Helmholtz had suggested that self-produced actions are critical for this learning, for they allow perceivers to test systematically the different possible causes of their sensory experiences (Helmholtz, 1867, vol. 3, ch. 26). Consistent with that suggestion, Held and Hein (1958) discovered that errors in hand-eye coordination only diminish when the prism wearer engages in active, self-guided movement. In their studies, people watched through a prism as their arm moved toward an object either passively (it was moved by an experimenter who operated a lever to which the arm was attached) or actively (the perceiver himself moved both his arm and the lever). The experiment was conducted with an ingenious apparatus that equated the visual stimulation and the motion in these two conditions, varying only whether the motion was self-produced. Adaptation occurred only for the participants who actively produced the movement. These findings provide evidence that purely associative learning about concurrent visual and tactile sensations does not underlie adults' flexible adaptation to visuomotor arrangements. When their visual experience serves as feedback about an attempted, self-guided action, however, perceivers can adjust to new visuomotor relationships.

Next, Held and Hein (1963) turned their attention to newborn kittens and conducted a landmark controlled-rearing experiment using another ingenious invention: the "kitten carousel" (figure 1.5). Like Gibson's kittens, the kittens in this study were reared in the dark, except for a period each day when they received visual experience in a device that propelled a pair of kittens through the actions of just one of them. Both kittens received the same visual information in

without differential reinforcement, how is the eventual behavioral preference acquired? Despite the different early experience, their preference for the shallow side gradually grew to that of the normal animal. . . . If this is learning, it does not fit any current definition." Sixty years later, their reasoning stands.

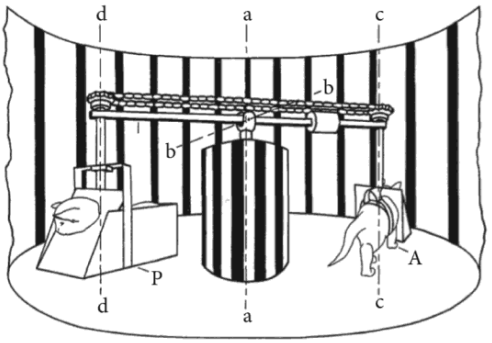

Figure 1.5 The kitten carousel (Held & Hein, 1963). Pairs of dark-reared cats experienced the same lighted environment together. In each pair, the "active kitten" (A) walked in a circle without sight of the ground or its own body. The "passive kitten" (P) sat in a carriage, propelled by A's movements. Thus, both cats received the same visual experience, produced by the same set of steps. Only for A, however, was vision correlated with self-produced motion. When tested on the visual cliff, active but not passive kittens avoided the deep side, despite their equivalent visual experience and lack of exposure to drop-offs. (Reproduced from Held & Hein, 1963.)

the carousel; both kittens were also equally free to move around in the dark. Only the active kitten, however, could experience the visual consequences of its own actions.

When the kittens were tested individually on the visual cliff, those who had moved actively in the carousel avoided the drop-off. Because these cats' visual experience occurred only while they were safely supported within the carousel, under conditions that blocked their view of their feet and the surrounding ground, this finding provides further evidence for Gibson and Walk's finding that cliff avoidance does not depend on experiences of simultaneously looking and walking onto surfaces varying in distance. The kittens who had moved passively in the carousel, however, did *not* avoid the deep side of the cliff. This finding sheds light on the experience effects that Gibson had discovered in cats: Visual experience during active motion, even without sight of the moving limbs, underlies kittens' capacity for visually guided locomotion.

With these studies, Held and Hein created a further bridge between psychophysical research on human adults and infant cats. Although kittens cannot conduct their own psychophysical experiments, Held and Hein applied the logic of Helmholtz's research to them and discovered common signatures in the patterns of adaptation shown by the two species. These signatures provide evidence that the mechanisms that produce adaptation in human adults also

produce adaptation in inexperienced animals. Could the existence of common mechanisms across animals be used to resolve the nativist-empiricist debate over the origins of space perception?

Two problems remained. First, none of these studies addressed the development of space perception in human infants, who are bathed in visual experience for 7 months before they begin to locomote independently. Second, none of these studies tackled the question Helmholtz had declared to be unanswerable: Do the animals in visual cliff experiments locomote preferentially on the shallow side of the cliff because they perceive the differing distances of the surfaces on the two sides, or because they detect differing patterns of sensation at their eyes? Perhaps animals only sense moving images, detecting a big change in image motions when they look from the centerboard to the deep side and a smaller change when they look from the centerboard to the shallow side. Animals may have evolved a connection between these 2D sensory patterns and locomotion: Newborn goats and dark-reared cats might avoid drop-offs *without* perceiving depth.

Held and Gibson each addressed this possibility through research with human infants. Interestingly, both began to study prelocomotor human infants in the 1970s, and both tested for signatures of mature human space perception in human infants. Because they investigated infants who were too young to reach for objects or to crawl toward them, both experimenters observed infants' systematic patterns of visual attention to varying spatial arrays. In some ways, their perspectives differed, but here I focus on their common ground.

Working in the tradition of open-air psychophysics, Gibson and her students asked whether the visual information that guides adults who judge distances across an open field—texture gradients and motion perspective—also guides infants who cannot yet reach for or crawl to objects. For these experiments, they used a variant on the looking time method of Fantz (1961). Fantz and others had discovered that when infants show no intrinsic preference for one visual array over another, their perception of a difference between the arrays can be tested by presenting one of the two arrays repeatedly before the critical preference test. Like adults, infants tend to look longer at novel arrays than at familiar ones, so this method reveals whether infants can discriminate between two patterns that they find equally attractive.

In a series of experiments, Gibson and others presented infants with patterns of texture and motion that lead adults to perceive surface rigidity and slant (von Fieandt & Gibson, 1959). They measured infants' looking time to these textures as the information was altered in ways that adults either do or do not perceive as a change in a surface's 3D properties. By testing all of these changes, they were able to determine which of them elicited an increase in infants' visual attention to the displays, and to compare infants' performance to adults' reported perception of changes in the surface layout.

This research revealed that infants showed all the signatures of adults' perception of surface rigidity (Gibson, Owsley, & Johnston, 1978/1991; Gibson et al., 1979/1991), slant (Day & McKenzie, 1973; Slater & Morison, 1985), approach (Ball & Tronick, 1971; Yonas, Pettersen, & Lockman, 1979; Carroll & Gibson, 1981), and occlusion (Carroll & Gibson, 1981; Granrud et al., 1984; Craton & Yonas, 1988). Infants increased their looking time only when shown visual changes that specify, for adults, a change in the 3D surface layout. In contrast, when infants were presented in succession with two distinct texture arrangements or motions that adults perceive as the same surface layout, they showed no attention to the change and treated the physically different patterns as equivalent.

Most impressively, 1-month-old infants who were familiarized with a rigid or flexible object by touch (exploring an object placed in their hand or mouth) subsequently looked longer at an object that they could see, but not feel, whose movement specified a change in substance: from rigid to flexible or the reverse (Gibson & Walker, 1984/1991). Like adults, infants perceived these displays not as a set of diverse, depthless, changing visual textures and motions but as a stable layout of surfaces in three dimensions: surfaces that can be touched as well as seen.

Albert Yonas, once a student of Gibson, complemented this work with studies of older infants' reaching for objects. When infants begin to reach for and pick up objects in the fifth month, they reach preferentially for things that are closer to them, providing further evidence for visual depth perception. As soon as reaching begins, Yonas found, it is guided by motion-carried information for relative surface distance, including the optic flow revealed by Gibson's psychophysical studies and the patterns of texture accretion and deletion that occur as observers move their heads while looking at a nearer textured surface that partly hides a farther one. At the start of reaching, infants also perceive depth from the binocular information given by convergence and stereopsis, to which I turn next. In contrast, infants began to use pictorial depth cues such as relative size to guide their reaching only later, by 6 or 7 months of age (Yonas, Granrud, & Pettersen, 1985). Early developing abilities to perceive distance from motion and binocular information may support the development of sensitivity to pictorial information for depth.

Meanwhile, Held revisited Fantz's (1961) finding that young human infants look reliably longer at a sphere than at a flat disc. To investigate whether infants responded to a 3D shape or to a complex pattern of depthless, changing sensations, Held and his collaborators focused on the signature limits of stereoscopic depth perception, revealed by a century of psychophysical experiments. When adults with normal binocular vision look directly at an object, the positions of edges, projected from the object to the two eyes, differ by small amounts, and

the angular sizes of these "binocular disparities" provide information for depth. Perception of stereoscopic depth is characterized by three signatures, all reflecting the geometry of binocular vision. First, we perceive one array in depth when binocular disparities are small but not when disparities are large: our eyes are too close together for objects to produce widely different images on our two retinas. Second, we perceive depth when small binocular disparities are horizontal but not when they are vertical, because our eyes are horizontally separated but vertically aligned. Finally, and most obviously, perception of depth from flat images with edges varying in binocular disparity only occurs when one wears stereoscopic goggles that project each image to one eye. If we look at the same arrays without goggles, we perceive images that are flat, fuzzy, and overlapping, rather than crisp surfaces arrayed in depth.

Held and his collaborators tested for these signatures in infants, using Fantz's method (e.g., Held, Birch, & Gwiazda, 1980). Infants were shown two arrays of vertical stripes through stereoscopic goggles, side by side. In one array, stripes were projected to the same positions at the two eyes; adults see this display as flat. In the other array, the edges of the stripes were projected to slightly different lateral positions; adults see these stripes as varying in depth (figure 1.6, left). Infants were tested between 10 and 40 weeks of age. Some infants at some ages also were tested with 90-degree rotated versions of these two arrays, producing vertical rather than horizontal disparity. In other control conditions, infants viewed the same displays but without the goggles. If infants perceive overlapping images rather than stereoscopic depth, they should look longer at *all* of the arrays presenting two different images, with or without goggles. In contrast, if infants perceive depth from stereopsis, and if this perception depends on the same mechanisms as in adults, then they should look longer at the arrays with differing images only when the binocular disparities are small and horizontal. Infants' preference for apparently 3D arrays should show the same signature limits as adults' perception of depth from stereopsis.

Held's experiments yielded three landmark findings (figure 1.6, right). First, infants showed no preference at all between the two arrays in any condition until the third or fourth month: the youngest infants showed no signs of preferring the image differences that lead adults to perceive either overlapping images or stereoscopic depth. These findings cast doubt on the claim that infants learn to perceive depth by associating tactile sensations with the overlapping image patterns (since the images aren't detectable during the critical months that precede infants' depth-dependent actions), and they reveal that stereoscopic depth perception emerges after birth. Second, infants' sensitivity to the disparate images grew very rapidly thereafter. Once infants began to detect image differences, their sensitivity to smaller and smaller differences grew so rapidly that within a week of detecting any differences, most infants were detecting the smallest disparities

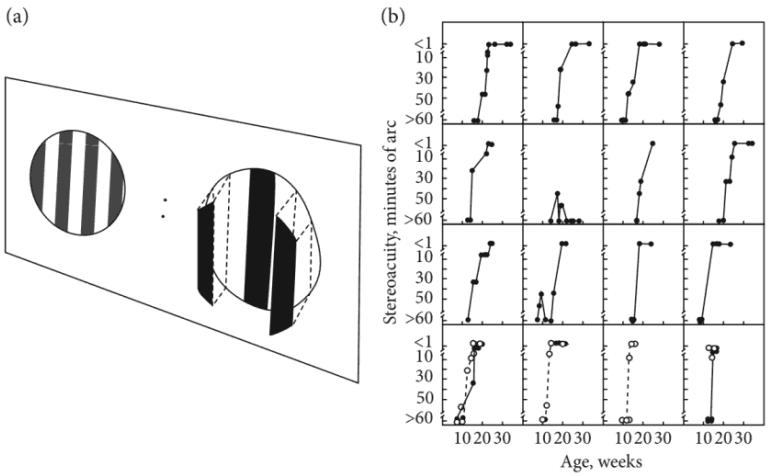

Figure 1.6 Displays (a) and findings (b) from Held, Birch and Gwiazda's (1980) tests of infants' perception of depth from stereopsis. The sample display is depicted as it would appear to an adult observing it through stereoscopic goggles: Perception of the flat figure on the left occurs when two identical images are projected to the two eyes; perception of bars varying in depth occurs when the two images present bars whose edges differ in their horizontal disparities. The data in (b) show the smallest horizontal disparity detected by each infant tested at weekly intervals, beginning at 10 weeks. All but one infant responded to small disparities by 3–4 months of age. (Images reproduced and figure adapted from Held et al., 1980.)

that Held's equipment allowed him to present, consistent with the high sensitivity to these disparities shown by adults. Third, and most important, infants' preference for the arrays with differing images showed the same signature limits as adults' perception of stereoscopic depth. Infants looked preferentially only at displays that looked 3D to adults: displays with image differences that were small and horizontal, viewed through stereoscopic goggles.

The story of the development of depth perception is not over, but let us enjoy this moment, and see where Held's and Gibson's experiments have led. Helmholtz articulated a method for discovering the mechanisms of space perception in human adults through systematic comparisons of the findings of experiments conducted by a community of psychophysicists. Although the perceptual experiences of each scientist are directly accessible only to that scientist, converging evidence from different scientists can provide evidence that the underlying mechanisms governing the perceptions of different scientists are the same. Of course, we can never be certain that the phenomenal experience of another person is the same as ours: We may both report that a flash of light is red

while experiencing different colors. Nevertheless, the psychophysical methods Helmholtz used allow us to predict what other perceivers will and will not report as surely as we can predict what our own reports of our perceptual experiences will be, and they reveal that the experiences of different perceivers depend on shared physical mechanisms.

Held and Gibson showed that the same logic applies to research on infants and animals. Just as we cannot know directly what other adults experience, we cannot know directly what infants or animals experience. We can, however, observe the behavior of infants and animals and use psychophysical methods to probe the signature properties of the mechanisms that guide this behavior. If the psychophysical functions that govern our own experience of depth also govern the locomotor patterns of cats on the visual cliff and the looking patterns of infants wearing stereoscopic glasses, then we can apply our psychophysics to them as securely as we apply it to other adults. The experiments of Gibson and Held placed the study of depth perception in human infants and animals on ground that is as firm, in principle, as the ground supporting psychophysical studies of space perception in adults.

Because infants younger than 4 months of age show no ability either to perceive stereoscopic depth or to attend to the double images by which an empiricist would expect infants to learn to perceive depth, Held's findings with younger infants do not speak to the role of experience in the development of stereopsis. Do infants learn to perceive stereoscopic depth over the first 4 months, perhaps by attending to double images under conditions that Held failed to elicit? Or do infants perceive stereoscopic depth innately but express that ability only at 4 months, either because the mechanisms of binocular vision mature only at that age or because binocular vision requires sharper spatial input than younger infants' visual system allows? One consideration suggests that the latter possibilities are more likely: At no age and in no study did young infants show preferences for displays presenting overlapping or binocularly disparate images that do *not* elicit perceptions of depth for adults. If infants learn that only horizontally disparate images specify changes in surface depth, then there likely would be some point in time at which an infant detects disparate images at other orientations, for how else would they learn which of these disparities are associated with different distances? Nevertheless, Held's studies leave open the question whether visually inexperienced human infants perceive depth on first encounters with an extended visual layout.

To address this question, investigators have focused on infants' perception of displays whose depth was specified by cues other than binocular disparity: especially the motion information studied by Gibson or the oculomotor cue of convergence discussed by Descartes and Berkeley. In the late 1980s, experiments in two laboratories provided evidence for depth perception in newborn human infants (Granrud, 1987; Slater, Mattock, & Brown, 1990). One of

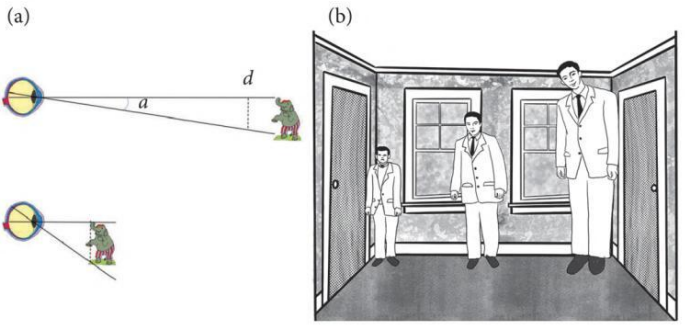

Figure 1.7 (a, top) Given the distance *d* of an object and the angle *a* that it subtends in the visual field, our visual system solves for the height of the object (dotted line). If the object moves closer to us, and the changes in its distance and angle are registered correctly, the perceived size of the object will not change (a, bottom). (b) This highly asymmetrical room is designed to give the false impression that it is symmetrical and that the two people therefore are equally far from us. As a consequence, the person on the right appears to be much larger than his companions at the center and the right corner of the room.

these experiments took a uniquely Cartesian (and Helmholtzian) turn: To determine whether newborn infants perceive depth, Slater and his collaborators asked whether the infants would use distance information to infer an object's size (figure 1.7a).

Compelling demonstrations that we as adults use distance in perceiving object size comes from photographs, taken in wildly asymmetric rooms, that lead us to misperceive how far away objects are from the camera (see figure 1.7b), as well as from systematic experiments in which the information about an object's distance is altered. When adults view an object through lenses that change the angle at which the eyes converge on it, thereby altering the information for the object's distance while leaving the object's angular size unchanged, we perceive a corresponding change in the object's size: If the changed convergence angle specifies that the object is more distant, yet its angular size does not diminish, we perceive the object to grow larger (Emmert, 1881; Helson, 1936; Boring, 1940; Wallach & Frey, 1972). Information for depth therefore influences perception of object size.

Slater and colleagues (1990) tested whether newborn infants' perception of an object's size is affected in the same way by changes in its distance. Using the novelty-detection method of Fantz and others, they tested infants with either a large object or a small object that were identical in their shapes and patterns but were presented at different distances so that they occupied same-sized regions of an infant's visual field (figure 1.8a). Prior to this test, half the infants

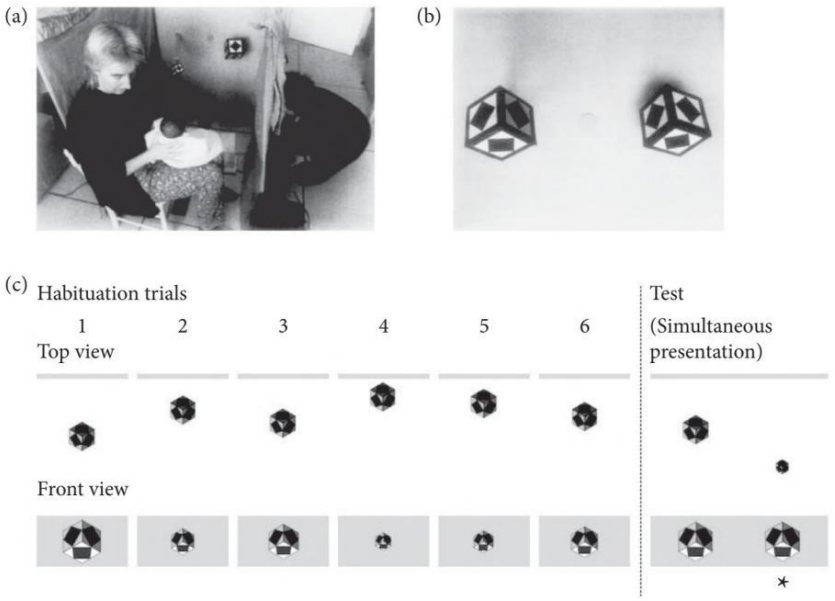

Figure 1.8 Experimental setup for Slater et al.'s (1990) tests of newborn infants' perception of angular size and distance. (a) The testing environment. The adult who holds the infant faces away from the display and does not look at the infant. (b) This photograph, taken from the infants' viewing position, shows the small cube on the left at a distance of 30.5 cm, and the large cube on the right at a distance of 61 cm. (c) After infants were familiarized with one of the two cubes, presented at a range of different distances (top view) and projecting a range of angular sizes (front view), infants looked longer at the cube with the novel true size (asterisk). (Images reproduced and figure adapted from Slater et al., 1990.)

were familiarized with the large object and half with the small object. On each familiarization trial, each object appeared at a wide range of distances (figure 1.8c, top), so that the infants in both conditions were exposed to images of a wide and overlapping range of sizes that included the image sizes presented at test (figure 1.8c, bottom). If infants perceived only the sizes of the images—not the distances of the objects or their actual sizes—they should show no preference between these test displays. Contrary to this prediction, infants looked longer at the object with the novel true size: Those who were familiarized with the larger object looked longer at the smaller one, and the reverse. During familiarization, infants evidently used information about angular size and distance to compute the sizes of the objects. This computation, in turn, provides evidence that the infants perceived the objects' distances, and that they performed the same geometric inferences that we make, as adults.

A wealth of further experiments, by Slater and others, showed that newborn infants perceive not only the constant sizes of objects over changes in the objects' distance but also the constant shapes of objects over changes in their orientation (Slater & Morison, 1985). Moreover, studies focused on infants' tracking of objects, conducted with infants as young as 3 weeks of age, provide evidence that infants perceive the constant position of a stationary object when the infants themselves are in motion. When a moving infant attends to a stationary object, she maintains fixation on the object by making highly accurate and coordinated compensatory movements of her head and eyes as the image of the object moves across her visual field. In contrast, when the same infant attends to a moving object while she is stationary, she is unable to perform the same smooth movements, even though the changes in her visual field are so nearly identical in the two conditions that adults, presented with films made by a camera placed in the position of the infant, are unable to discriminate them (Rosander & von Hofsten, 2000). The ability to distinguish motion of the self from motion of the world, perceiving the constant positions of objects as we move, has deep roots in human development.

These findings settle two issues in the debate between Descartes and Berkeley, in Descartes's favor. First, inexperienced human infants do not appear to experience a confusion of depthless sensations but a spatially coherent world. Second, infants gain knowledge about the world not only through the shaping effects of their experiences acting on the environment but through processes akin to rational inference: They infer not only the distance of an object from the two directions in which their eyes are pointed inward but also the size, position, and motion of an object from other geometric variables, including the size, shape, or position of the object's image in the infant's visual field, and the distance, slant, or motion of the infant relative to the object.

In summary, the research described in this section effectively answers an ancient question: It shows that humans are endowed with an innate capacity to perceive the spatial structure of the surrounding visual layout. This research testifies to the existence and importance of innate perceptual capacities, but it also speaks to the nature and importance of learning. The prism adaptation experiments of Bedford and Gharamani suggest that infants' early capacities for visual space perception support a continuous learning process, by which children adjust their perception of the positions of their limbs to the growth of their bodies (if the mechanism of adaptation that Bedford found in adults also operates in infants). Similarly, Held's studies of perceptual adaptation in human adults and infant kittens sheds light on the role of self-produced motions in the development of visually guided action, and Yonas's studies of the emergence of infants' sensitivity to pictorial depth information suggest how infants build on their early-emerging abilities to perceive depth from stereopsis, convergence, and optic flow

to learn further cues to depth, such as the patterns of converging lines used by Renaissance painters. Infants' innate capacity to perceive a spatially organized, 3D world likely serves, first and foremost, as a basis for learning more about the world's perceptible properties.

Capacities to perceive the spatial layout also serve to support infants' learning about the consequences of their actions. When infants begin to crawl, for example, Gibson and Walk's studies reveal innate capacities for depth perception on the visual cliff, but they provide no evidence that infants understand that cliffs are dangerous or that their bodies are vulnerable to injury. Indeed, Gibson's original observations suggested that infants lack this understanding: Newly locomotor infants showed no fear as they gazed from the cliff's center board at the distant visible surface (E. J. Gibson, 1991) and instead showed interest as they gazed over the cliff (Campos et al., 1978). Consistent with this observation, infants presented with a cliff with no glass protection are quite apt to fall as they explore it by crawling (Adolph, 2000) or walking (Adolph, Berger, & Leo, 2011). The onset of fear of heights is related to the onset of independent locomotion and almost certainly depends on learning from infants' own self-produced motions (Campos, Bertenthal, & Kermoian, 1992; see Adolph, Kretch, & LoBue, 2014 for review), like the learning of kittens on Held and Hein's carousel.

Beyond the development of fear, research by Karen Adolph, a former student of Ulric Neisser and Gibson, reveals a prolonged and fascinating development of capacities for coordinating surface perception with action in infants and young children who locomote on slopes, bridges, and other variations of the original visual cliff. Although infants perceive surfaces from the first months, they discover the affordances of those surfaces for their own developing actions gradually and progressively, as new action capacities emerge (Adolph, 1997; see Adolph, 2008, for a review). Infants' crawling, cruising, and walking are well adapted to some visible surface properties, especially surface distance and rigidity, but not to other surface properties, such as slant (Adolph, Eppler, & Gibson, 1993; Adolph, 1995, 1997) or deformability (Joh & Adolph, 2006).

Children eventually learn to take account of these properties, but they must learn to adapt to the same properties again and again as new action capacities emerge. For example, infants who have learned not to descend a risky slope or step into a gap when crawling must learn again to avoid the same slope or gap when walking (Adolph, 1997, 2002; Adolph et al., 2011). Although human infants perceive an extended spatial layout from the beginning, they spend many months learning the affordances of this layout for their own developing actions. All these findings suggest that innate visual capacities support learning: Nativist and empiricist hypotheses are in dialogue, not debate, as their proponents jointly work to understand the power and effectiveness of vision.

4 Marr and the Modern Interdisciplinary Synthesis

While Gibson and Held were pursuing their psychophysical experiments in the 1950s, transformative developments were occurring in three fields. In neuroscience, Hubel and Wiesel (1959) published their first study of the response properties of individual neurons in the primary visual cortex of cats, launching the growth of knowledge of the detailed connectivity, function, and development of mammalian visual systems. In molecular biology, Watson and Crick (1953) worked out the structure of DNA, setting in motion a process that has led to the meteoric growth of molecular biology and genomics, providing neuroscientists with powerful tools for studying the activity of neural circuits in the visual system and elsewhere. And before the field of computer science had a name, Alan Turing, who had spent the World War II years building machines that solved complex problems in cryptography faster than humans, wrote "Computing Machines and Intelligence" (1950): an early harbinger of the now flourishing fields of machine learning and artificial intelligence.

In the late 1970s, the visionary scientist David Marr, who had been trained in mathematics and neuroscience, joined MIT's Department of Psychology (its name soon changed to Brain and Cognitive Sciences), as Held became its chair. In the few short years before his death in 1980, Marr revolutionized the field of computational vision with new findings, new hypotheses, a new articulation of the goal of vision itself, and new arguments for an interdisciplinary science of vision centered on experimental psychology, systems neuroscience, and computational modeling (Marr, 1982). I focus here on Marr's prescient ideas about the nature of explanation in cognitive science, the fundamental function of vision, and the content of the representations that the visual system forms. Then I turn briefly to more recent developments in the interdisciplinary visual sciences.

Marr's most-discussed contribution to cognitive science comes from his argument that the study of human vision should aim for explanations at three levels. Explanations at the lowest level ("implementation") focus on the neurons and connections underlying vision in animals or the silicon circuits underlying visual processing by machines. Explanations at the next level ("algorithm") focus on the steps by which the neural or artificial hardware performs tasks of processing information from light. Explanations at the highest level ("computational theory") specify the tasks performed by the visual system and its subsystems—that is, the functions that their algorithms implement—and the properties of the visual world that allow perceptual systems to accomplish these tasks. This is the level at which James and Eleanor Gibson focused their work, but Marr moved beyond them by harnessing the mathematics and engineering of computer science.

Working in this perspective, Marr made landmark, enduring contributions that have reverberated through cognitive science, as will be evident throughout this book. He argued, first, that visual representations and computations form a hierarchy. At the lowest level of processing, the visual system functions to detect changes over space and time in 2D arrays of light: *edges* and their displacements, at a variety of scales. At the highest level of processing, people build representations of the 3D shapes of objects. Marr noted, however, that sets of rules for parsing arrays of pixels into meaningful objects don't work well, because the boundaries of objects aren't clearly specified in the edges that meet the eye or camera in any particular 2D image, and because recognition of useful categories of objects requires specific knowledge of the properties and functions of those objects. Nevertheless, Marr conducted ground-breaking work on the representations of object shape that serve to categorize objects. I discuss this work in chapter 6.

Between edge detection and object recognition, Marr discerned an intermediate level of visual representation, the construction of which constituted the fundamental function of vision and the end, he suggested, of "pure perception" (Marr, 1982, p. 268). Marr called this representation the 2.5D sketch, as it fails to capture the full 3D structure of the environment but goes significantly beyond the 2D arrays that light-reflecting surfaces project to the eye of an observer or camera. This representation captures the distance and orientation of each detectable patch on each surface that is visible to an observer from a single station point, measured relative to the observer: how far away from the eye or camera each visible surface patch is, and how it is oriented.[9]

The 2.5D sketch represents these visible surfaces as a mesh: a complexly shaped array of surface patches, with abrupt discontinuities where one surface ends and the surface behind it enters the visual field. The geometric information captured by the mesh comes from earlier visual processes operating on representations of the 2D edges in the array, including processes for analyzing edges' patterns of motion, or for analyzing the small differences in the arrays of edges detected at the two eyes that give rise to perception of depth from stereo. Detailed study of these processes, and others, yielded basic insights into the properties of the visible surface layout that the visual system builds on. For example, the processes that construct the 2.5D sketch accord with the principle that surfaces change smoothly, in orientation and depth, almost everywhere in a visual scene. When

[9] I oversimplify: Marr noted, and subsequent research confirms, that observers are more sensitive to the relative distances and orientations of surface patches that are adjacent rather than separated. For example, we can see more accurately that a surface has a small bump, where points on the surface are closer to us than are its neighbors, than that one such bump is closer to us than another bump that is spatially separated from it. First and foremost, we represent visible surfaces as continuous almost everywhere, except at their edges where they occlude more distant surfaces (Gibson, Kaplan et al., 1969; Gibson, 1979; Tsao & Tsao, 2021).

the same scene is represented at multiple scales, from large surface patches that are coarse and blurry to tiny patches that are fine and sharp, the important edges in the scene occur where discontinuities at these different scales coincide.

Marr's theory of vision makes explicit the central insights from the Gibsons' research on outdoor psychophysics, and it grounds their insights in a computational theory. In the brief time he had to work on vision, Marr did not, to my knowledge, consider the problem of visual development, although the work of Gibson and of Marr's colleague, Held, suggested that a theory that captures our perception of surfaces as human adults would apply to infants as well. Four years after Marr's book appeared, however, a different computational approach to vision, begun many decades earlier, was heralded by an important collection of articles by David Rumelhart, James McClelland, Geoffrey Hinton, and others, published in two volumes, *Parallel Distributed Processing* (1986). The articles described their research with machines that were programmed to interact like a composition of elementary, interconnected units, resembling the multilayered neural networks in the visual system, with an input layer like the retina, a single output layer like the reports given by participants in a psychophysical experiment, and a set of intermediate layers like those found in the visual cortex. Instead of connecting these units selectively and spatiotopically, as in the brain, however, all units in one layer were connected to all units in the layers above and below it, and the starting strengths of individual connections between units in adjacent layers were set at random. Thus, the machines initially showed no coherent relation between the input and output layers. They were programmed to learn, however, by modifying the strengths of their connections, so as to approach progressively an initially specified function between input (say, a handwritten character) and output (say, a symbol for the letter *A*).

These models captured a number of interesting properties of visual categorization and have had a large and enduring impact on cognitive psychology and computer science. Although they initially failed to produce a great leap forward in computer vision, more recent research with neural networks has had better results, using machines with faster, more powerful processors, with access to big data from internet companies and platforms, and programmed to instantiate mathematical functions called convolutions to optimize detection of environmental features at different locations and scales, as did Marr's edge-detection algorithms. These machines have shown striking success at what, for 50 years, were intractable problems of recognizing handwritten characters, objects in natural images, or human speech (LeCun, Bengio, & Hinton, 2015).

Convolutional neural networks do not learn or generalize as humans do, however (Lake et al., 2017; Mitchell, 2019). Even the best current artificial neural networks make generalizations that no sentient adult or child would make, and they fail to make generalizations that come to children naturally. For example,

machines trained on images of airplanes in plausible scenes will respond similarly to airplanes in scenes that are highly implausible, unless trained not to do so. In contrast, children who have seen and ridden in real cars don't need to be taught, the first time they see a small plastic object with the shape and markings of a car, that the object is not a car but a toy that was designed to look like one (DeLoache, Uttal, & Rosengren, 2004).

Might artificial neural networks perform better if they learned to recognize objects not directly from pixels but via intermediate, viewer-centered representations of the 2.5D surface layout? Recent research on shape-based object recognition suggests that they will (e.g., Wu, Wang et al., 2017; see chapter 6). For this to be possible, however, the visual system must be able to compute the 2.5D sketch prior to the development of knowledge of any of the particular kinds of objects that those scenes contain. How might the 2.5D sketch be constructed by inexperienced human infants?

To consider this possibility, I make a brief detour into the fields that inspired the recent research on artificial neural networks, from systems and cognitive neuroscience. The work of Hubel and Wiesel, and of their many successors and descendants, supports three conclusions. First, the visual system shows exquisitely precise wiring at birth. At the start of visual experience, the visual cortex of cats and monkeys has the complete, hierarchically organized, and layered and topographic structure of the mature visual cortex. Individual neurons, though less stable in their firing than those of mature animals, show response patterns that are qualitatively similar to those of adult animals. Second, connections between lower and higher levels of the visual system proceed in both directions: there are at least as many connections from higher to lower levels as there are in the opposite direction. Third, the visual system responds adaptively to altered visual input. If one eye is covered, neurons that respond more to input from the open eye come to predominate. Newborn brains are wired both to see at birth and to adapt to changes in visual experience.

What allows for the initial, highly precise wiring of mammalian visual systems? Guided by developments in molecular genetics, research reveals that the primary structure and connectivity of the visual system is determined by a precise combination of genes that are activated in the right places and at the right times to build an appropriately connected central nervous system, independently of fetal activity or sensory experience. (For review, see Ackman & Crair, 2014.) The newborn brain also is active before birth, and in many decades of research, Carla Shatz, Larry Katz, Michael Crair, and others have studied this activity and its consequences. In fetal or dark-reared rodents and cats, the same activity-dependent processes that underlie perceptual learning from postnatal visual experience also occur prior to the onset of visual experience. Before birth, waves of spontaneous activity pass over the sheet of ganglion cells in the retina

that connect first to the subcortical structures in the thalamus and then on to the cortex. These activity patterns modulate the detailed wiring of neurons from the eye to the thalamus and cortex, even before the photoreceptors in the eye are mature enough to respond to light, and well before the developing animal encounters its first visible environment. By altering the strength of the synaptic connections that fetal neurons form, these activity patterns are thought to contribute to the detailed functional specificity of the visual system at birth (e.g., Katz & Shatz, 1996; Ackman, Burbridge, & Crair, 2012).

Does spontaneous activity from the retina to the primary visual cortex shape the mechanisms that allow Gibson's newborn goats and dark-reared rats to detect a visual cliff, or to construct Marr's 2.5D sketch, prior to any experience of a visible environment? These problems of vision are known to be extremely difficult, because many geometrically different visible surface layouts could be the source of any single edge that is detected by a neuron in the primary visual cortex. How might the visual system develop the connectivity that allows animals to perceive a spatially organized visible layout at the onset of visual experience?

One possibility comes from an insight that is as old as the Renaissance. Although it is hard to recover the 3D objects in a scene from the 2D images of the scene that are detected by the two retinas, the inverse problem is easier. If one starts with a given 3D (or 2.5D) representation of a visible scene, one can use basic rules of geometry (or the act of drawing on a window) to determine how to render that scene into a realistic landscape painting. In contemporary computer graphics, such rendering is performed by a graphics engine: a computer program that takes as input a description of the positions, orientations, shapes, lighting, and movements of the surfaces and objects in the scene, and outputs a succession of 2D images of the scene.

Forward graphics is a highly developed and successful field of computer science. It has generated animated films that are so lifelike that their designers now must work to reduce their fidelity, lest the animation be mistakenly seen as the more pedestrian product of a regular movie camera. With somewhat lower fidelity, graphics engines can generate these animations fast enough to serve in rapid-fire, interactive video games. In contrast, computer vision systems still struggle to attain the accuracy and speed that would allow for fully autonomous self-driving cars or robotic butlers. This struggle reflects the greater difficulty of inverse graphics: the design of computer programs that begin with 2D images and recover from them the positions, orientations, shapes, lighting, and movements of the objects and surfaces that are their most likely source. Inverse graphics is hard, because any single 2D image of a scene can be produced by many different combinations of these variables. Forward graphics is not easy, as the exciting recent history of animated films reveals, but it is easier than inverse

graphics, because each combination of surfaces, objects, light sources, and camera positions will give rise to a single 2D image.

These observations raise a possibility: In the fetal brains studied by Shatz, Katz, and others, might there be evolved patterns of connectivity that implement the easier operations of forward graphics: a neural graphics engine, generating activity that mimics the optic flow that inexperienced animals encounter on the visual cliff? If such a mechanism exists, it likely will be found in subcortical brain regions, because even insects are sensitive to optic flow. The activity that it generates might propagate, however, through the visual system, to regions of the visual cortex that represent the 3D or 2.5D surface layout, where descending patterns of spontaneous activity could interact with the bottom-up activity that is generated in waves by the retinal ganglion cells discovered by Shatz, Katz, and their colleagues. In this way, the top-down activity of an innate forward graphics engine might provide higher-level visual representations that serve as training targets for the neural system that begins in the retina and propagates activity upward through the visual system, preparing fetal animals for their first encounters with visual cliffs, obstacles, and traversable surfaces.

Research on human fetuses, using methods of functional brain imaging, suggests that such processes might occur before birth. This research reveals spontaneous activity not only in the sensory cortices but in the temporal cortex (the site of much of the visual information underlying object recognition), in the parietal cortex (the site of much of the visual information underlying actions such as reaching or navigating), and in the prefrontal cortex, where representations of places, objects, and people may together inform the generation of action plans. (See Dehaene-Lambertz & Spelke, 2015, for review.) All these areas connect to older, subcortical brain structures. Although the sources and functions of their prenatal activity are not known, the existence of such activity suggests that fetal brains could, in principle, be furnished with ancient, generative systems like graphics engines (Hinton et al., 1995; Ullman & Tenenbaum, 2020).

Could such systems shape the fetal visual system to perform the inverse operations that guide the locomotion of newborn goats onto traversable surfaces while preventing them from walking off the edge of a cliff, and that allow newborn human infants to perceive the constant sizes of objects as they change in distance? Recent research provides suggestive evidence that alternating cycles of synthesis, implemented in a symbolic, probabilistic computer program like a graphics engine, and of analysis, implemented in an artificial neural network, can account for the response properties of the three primary cortical areas that represent faces in monkey brains (Yildirim et al., 2020). Moreover, one such hybrid model has learned not only to classify objects from images by using a representation like Marr's 2.5D sketch but also to generate new images of the objects after changes to their positions, lighting, or other scene features (Wu, Wang et al.,

2017). I discuss research on face perception in chapter 8 and on object recognition in chapter 6.

Here, I end with a finding that emerged as this book was entering production, and that speaks more directly to the research presented in this chapter. Ge et al. (2021) studied visual activity in the brains of developing mice, after birth but before the onset of visual experience. (Mice spend their first two weeks with eyes closed.) With clever optical techniques, they imaged, both simultaneously and over time, the many pairs of neurons that synapsed in the superior colliculus: an ancient, subcortical structure in the visual system with homologues in a wide range of animals, including humans. The incoming neurons were retinal ganglion cells that collected information from each eye; the receiving neurons resided in the superior colliculus itself. From birth onward, waves of activity were recorded from the retinal ganglion cells, but these waves conformed to a special pattern at a particular developmental moment: 8 to 11 days after birth (and still before eye opening). At that time, the activity of the retinal ganglion cells, measured both at the retina and at the synapse in the superior colliculus, corresponded to the patterns of optic flow that animals experience when they move, with eyes open, through a lighted environment. Thus, the optic flow used by Gibson and Held's animals was simulated by these mice prior to the onset of visual experience.

For methodological reasons, these experiments were conducted on mice who were genetically engineered to fail to grow a cerebral cortex. Lacking any descending cortical connections, the retinal ganglion cells of these mice failed to form synapses with neurons in the thalamus (Shanks et al., 2016): a key way station to the visual cortex. The optic flow pattern that Ge et al. discovered therefore likely originated either in the two retinas or in the superior colliculus. Interestingly, the observed pattern of optic flow was the same throughout this four-day period, but the cell types and synaptic connections that produced it in the retina changed between days 9 and 10. Is this pattern preserved over these retinal changes because it is produced by descending signals from neurons in the superior colliculus that function as a rudimentary forward graphics engine? Or is the pattern produced within the retina itself: the hypothesis the authors favor? In either case, waves of activity, simulating the optic flow that animals encounter when they begin to move through visible layouts, propagate from the retina to the superior colliculus of animals who have yet to open their eyes. By some mechanism, visually inexperienced animals dream of moving through visible landscapes. Such dreams may prepare Gibson's dark-reared rats and newborn goats for interpreting the visual information they receive as they stand on the center board of the visual cliff.

In summary, research in computer graphics and computer vision, building on research in psychology and neuroscience, suggests how innate capacities for

visual space perception might arise in visually inexperienced animals. Activity-dependent learning processes are widely believed to underlie the adaptability of the visual system to its own growth and postnatal experience, but these processes may do more: They may contribute to the emergence of abilities that animals display at the time of their first encounters with a visible environment. By generating activity that propagates through the visual system, these processes may create, in the brains of such animals, systems for perceiving the visible surface layout through which they will come to move.

More generally, the research discussed in this chapter, culled from experiments in psychophysics, visual neuroscience, and artificial intelligence, suggests that innate knowledge and learning are deeply intertwined, and that both pose hard problems but no imponderable mysteries. Converging scientific developments have transformed claims of nativism and empiricism from topics of debate in philosophy into a vibrant interdisciplinary science of vision. In later chapters, I suggest how such activity may contribute to our commonsense understanding of objects, places, and people, as we simulate object collisions, imagine paths through navigable environments, or infer other people's action plans.

5 Looking Ahead

I have begun this book by writing about a phenomenon that lies outside its purview. The capacity to perceive the visible surface layout is not a product of any system of core knowledge: It gives rise to experiences of the world as we sense it, not as we know it to be. Here I ask how the methods and findings that shed light on the origins of this capacity might be leveraged for studies of the nature and origins of human knowledge.

The system that underlies our visual perception of space has three key properties that made possible all the discoveries reviewed in this chapter. First, visual space perception has a dedicated system of inputs: it is impenetrable by knowledge obtained in other ways. Studies of visual space perception therefore avoid the complexities that arise when mental states can be influenced by everything that a person believes, and they allow for a systematic psychophysics that relates our visual perceptions to a limited set of stimulus parameters. Second, space perception begins with abilities that remain present and functional throughout our lives. Vision therefore is not wholly a product of human history and culture; it has a constant core. Third, many aspects of the human visual system are shared by other animals, and so animal models allow for research using methods that are not ethically or practically possible with humans, including controlled rearing, invasive neuroscience, and experiments in genomics. Relatively simple animals like flies exhibit basic visual processes without a massive overlay of

knowledge about other properties of the world, making them good targets for computational models of vision.

Later chapters will reveal that core knowledge systems have the same three properties: They are relatively encapsulated and respond to a limited set of inputs, they begin to function early in development and persist throughout our lives, and they are shared by many different animals, as their structure and functioning were shaped, over hundreds of millions of years of evolution, to capture enduring properties of the things we see. Thus, the systems at the center of this book can be studied by means of the panoply of methods in experimental psychology, neuroscience, and computer science that have figured in this chapter. For example, powerful insights into core knowledge systems have come from the discovery of signatures of each core system, like the signatures of stereoscopic depth perception studied by Held. The methods and insights from vision science will be important guides as we attempt to disentangle the multifaceted, interconnected capacities underlying children's developing knowledge.

The last reason that I begin with a chapter on vision is opposite to the other reasons: The material in this chapter will allow me to clarify, in later chapters, what core knowledge systems are *not*. Although core knowledge systems have much in common with the visual system, they differ from all perceptual systems in crucial ways. Like commonsense and formal theories, core knowledge systems center on interconnected, abstract concepts that organize our intuitive understanding of the world. In contrast to our perceptual systems, core knowledge systems do not solve the problem of building a representation of the spatially extended, perceptible surface layout from incoming sensory information. They focus instead on the problem of understanding what the sensed world consists of: what entities inhabit it, how those entities behave, and why they do what they do.

Are there cognitive systems, beyond perception, that bring us knowledge of how the world works, but that share the three key properties of the visual system that I just outlined? Studies of infants provide evidence that there are. Research on core knowledge challenges the pervasive view, in philosophy and psychology, that cognition subdivides into automatic, implicit perceptual systems and deliberate, conscious systems of thought, with no territory in between. The rest of this book aims to chart some of the territory that lies between, and connects, perceiving to thinking.

2
Objects

Objects are the primary things that we perceive, act on, categorize, name, count, and track over time. Like the perception of surfaces in depth, the experience of unitary, bounded, solid objects arises immediately and effortlessly. Even in the simplest scenes, however, our experience of objects extends beyond what we directly see or feel. We experience objects as standing in front of a background that continues behind them, even in pictures (Rubin, 1921/1958), and as complete, solid bodies, although their backs are hidden. When the fronts of objects are partly hidden behind other objects, we can more easily describe or draw their complete shapes than their visible surfaces. When objects move fully out of view, we sense their presence and extrapolate their motion (Michotte, Thinès, & Crabbé, 1964).

Representations of objects pervade our language and commonsense reasoning. We categorize and name objects so readily that a line drawing of a chair will call its name to mind, unconsciously and subvocally, even when it is irrelevant to our present concerns (Meyer et al., 2007). Presented with an array of stationary objects, we detect whether some are perched more precariously than others (Battaglia, Hamrick, & Tenenbaum, 2013). When two objects collide, we expect them not to interpenetrate, and we infer some of the forces that each exerts on the other (Michotte, 1963; Lesliea, 1984a; Kominsky et al., 2017). Finally, when we view a set of objects, such as a flock of birds or a plate of oranges, we gain an immediate sense of approximately how many objects are present without counting, and we know exactly how many objects are present when numbers are small (Feigenson et al., 2004). Object cognition therefore spans a host of phenomena, from the detection of visible bodies to the apprehension of abstract forces and number.

Object cognition is surprisingly opaque. We often have vivid and compelling intuitions about the conditions under which objects persist or perish, but attempts to codify those intuitions invite counterexamples and hard questions. Are you the owner of the same car, if an accident required that you replace most of its parts (Hobbes, 1655/1839–1845)? If your child rolls her Play-Doh elephant into a ball, is it the same toy as before (Hume, 1739/1962; Wiggins, 1980)? Names for objects are among children's earliest words, and people all over the world single out the same sorts of things as nameable objects: we have words for the cow and for the milk pail, but no single word for the udders, pails, and bottles

What Babies Know. Elizabeth S. Spelke, Oxford University Press. © Elizabeth S. Spelke 2022.
DOI: 10.1093/oso/9780190618247.003.0002

that contain milk, or for the physical arrangement of cow, pail, and stool in which milking occurs. Nevertheless, attempts to state rules for singling out nameable objects have a long and difficult history in philosophy and psychology (Fodor et al., 1980; Kripke, 1972).

Despite these difficulties, cognitive science has gained considerable understanding of our capacity and penchant for representing objects, thanks in part to insights gained from human infants. Studies of object cognition in infancy suggest that an interesting subset of the diverse phenomena of mature object cognition—including aspects of object perception and of intuitive physical reasoning—are products of a unitary system for singling out objects and making sense of their behavior. Like the visual system, this system is at least partly innate, it persists over later development, and it operates automatically and effortlessly: When we look at an array, we cannot avoid seeing the objects it contains. In contrast to any perceptual system, it centers on a set of concepts that are both abstract and interconnected, and it allows us to reason and learn about objects and their interactions not only when we perceive them but when we think about them. These properties of the core object system—its unity, its persistence over development, its support for our reasoning and learning, and its place within our cognitive architecture, beyond perception but short of fully deliberate thought—apply, I suggest, to all our systems of core knowledge.

In the first half of this chapter, I discuss a body of findings providing evidence that young infants perceive (section 1), track (section 2), and reason about objects (section 3) before they can effectively manipulate them. Then I turn to research probing the source of these abilities and present evidence that they depend on a single cognitive system (section 4). The system is not a temporary scaffold but an enduring foundation both for our knowledge and reasoning about objects as adults and for children's learning about objects and their mechanical interactions (section 5). I ask next whether core knowledge of objects is innate, reviewing research that makes this claim plausible but supports no broad and general conclusions (section 6). In the concluding summary, I propose that an early emerging, core system of object representation is distinct from, and interposed between, our perceptual systems and our systems of explicit beliefs (section 7).

1 Object Perception in Human Infants

When I started puzzling over the nature and development of object perception, in the late 1970s, few psychologists thought that infants either perceive or reason about objects. Most experimental psychologists believed, with William James (1890), that perception of arrays of visible objects begins with a confusion of sensations, elicited by the complex and changing mosaic of surfaces reflecting