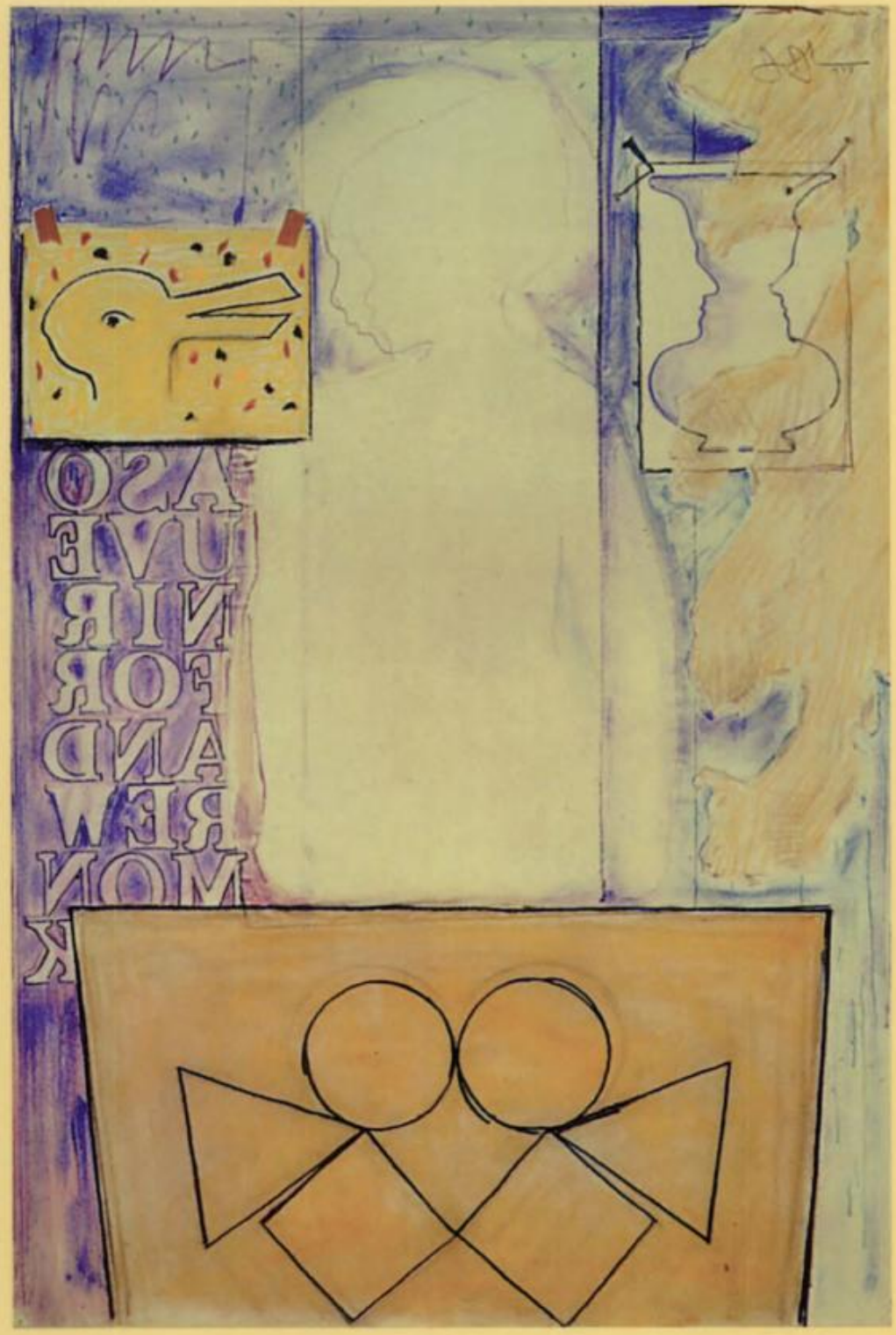

Words, Thoughts, and Theories

Alison Gopnik and Andrew N. Meltzoff

Words, Thoughts, and Theories

Alison Gopnik and Andrew N. Meltzoff

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

This One

First MIT Press paperback edition, 1998

This book was set in Bembo by Asco Trade Typesetting Ltd., Hong Kong, and was printed and bound in the United States of America.

Third printing, 2002

Library of Congress Cataloging-in-Publication Data

Gopnik, Alison.
Words, thoughts, and theories / Alison Gopnik and Andrew N. Meltzoff.
p. cm. — (Learning, development, and conceptual change)
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-07175-4 (hc : alk. paper), 0-262-57126-9 (pb)
1. Cognition. 2. Cognition in children. 3. Psycholinguistics. 4. Philosophy and cognitive science. I. Meltzoff, Andrew N. II. Title. III. Series.
BF311.G63 1996
153—dc20 96-11804
CIP

10 9 8 7 6 5 4

Contents

Series Foreword

This series in learning, development, and conceptual change will include state-of-the-art reference works, seminal book-length monographs, and texts on the development of concepts and mental structures. It will span learning in all domains of knowledge, from syntax to geometry to the social world, and will be concerned with all phases of development, from infancy through adulthood.

The series intends to engage such fundamental questions as the following:

The nature and limits of learning and maturation The influence of the environment, of initial structures, and of maturational changes in the nervous system on human development; learnability theory; the problem of induction; domain-specific constraints on development

The nature of conceptual change Conceptual organization and conceptual change in child development, in the acquisition of expertise, and in the history of science

Lila Gleitman
Susan Carey
Elissa Newport
Elizabeth Spelke

Preface and Acknowledgments

Almost exactly twenty years ago, two extremely young and rather arrogant graduate students had their first lunch together in an Oxford dining hall. They remember very little about the food, though it was almost certainly terrible, or the surroundings, though they were almost certainly beautiful. They do both vividly remember the conversation, though. The younger (and somewhat more arrogant) of the two, who in those days was an evangelical Chomskyan, said she didn't think there was much point in reading Piaget, and the older (and somewhat wiser) replied that he thought actually Piaget was pretty interesting. The argument went on until dinnertime. The conversation they began that day has gone on ever since, across five cities, three countries, and two continents; by letter, phone, e-mail, and fax; to the successive accompaniment of babies cooing, toddlers pretending, schoolchildren multiplying, and teenagers playing hallway roller hockey. They are a bit grayer and more lined, and some of the sharp edges have been knocked off, but their youthful pleasure in just talking about this stuff has never diminished. This book is the latest installment of the talk, and we hope it will give other readers and talkers some of the same pleasure.

As we have grown older the conversation has widened to include colleagues, friends, teachers, and students around the globe, and we want to thank all these common and individual interlocutors. Our first Oxford conversation would never have taken place without Jerome Bruner. In fact, for fifty years or more he has been stimulating this kind of talk, and our lives, as well as the whole large conversation that is cognitive science, would not have been the same without him. This particular turn in the talk has also depended on Harry and Betty Stanton, who invited us to

write for the MIT Press, and to Amy Pierce, our editor there. A number of our colleagues and friends read and commented on various drafts of the manuscript, and we are very grateful for their attention and time. They include Dan Slobin, John Heil, Carolyn Mervis, Simon Baron-Cohen, and Daniel Povinelli. Susan Gelman, John Campbell, and Liz Spelke also served as exceptionally acute, sensitive, and generous reviewers for the MIT Press—all authors should have such reviewers.

At this point we need to acknowledge our individual debts too, so Alison is now speaking rather than the authorial we. (Managing the first person has been an interesting challenge throughout this book.) I am exceptionally fortunate to spend my days in the psychology department at Berkeley. The exciting talk wafting through the corridors of Tolman Hall is the product of all its individual minds, but I particularly want to acknowledge my conversations with Dan Slobin, John Watson, Steve Palmer, Lucy Jacobs, and the late Irvin Rock, who have all contributed specifically to this book. Like many others, I also owe a special moral debt to Irv Rock for demonstrating how goodness and intelligence can be combined. My students here, I am happy to say, are just as great talkers as we were at Oxford, and the book has benefited from conversations with all of them, particularly Virginia Slaughter, Betty Repacholi, Therese Baumberger, Jennifer Esterly, Reyna Proman, Andrea Rosati, and Eric Schwitzgebel. The support of the National Science Foundation, grant no. DBS9213959, has been crucial.

Henry Wellman suggested that we write a chapter on the theory theory together, and that joint work was the source of much of the theoretical material in chapter 2. Rarely has writing a chapter been so fruitful! His comments and discussion have always been illuminating. I have tried to move back and forth between psychology and philosophy. Two philosophical mentors have helped enormously in allaying the anxieties of that amphibious project (I leave it to the reader to decide which is ocean and which dry land). John Campbell has constantly been a thoughtful and imaginative reader and correspondent, with an inimitable knack for coming up with precisely the point I was trying to get at myself and then coming up with brand new points I would never have gotten myself but that seem exactly right. Clark Glymour discussed almost everything in this book at one time or another by e-mail. He also read the sections on philosophy and science and gave me a terribly hard time about them. This is a much better book as a result and comes much

closer to emulating his own intellectual seriousness and rigor, and I am grateful.

I spent my own childhood in constant conversation in the best setting for intellectual research I have ever known. My warmest gratitude and love goes to the directors of the Gopnik Academy, Irwin and Myrna, and to my colleagues there, Adam, Morgan, Hilary, Blake, and Melissa. It's conventional at this point to apologize to your children for neglecting them while you got your book written and to express your gratitude for the fact that they absented themselves during the process. The apology is in order, but for this particular book and these particular children, Alexei, Nicholas, and Andres Gopnik-Lewinski, the acknowledgement has to be a bit different. Without their constant presence, their conversation and company, without the example of their insatiable curiosity about the world and their genius in figuring it all out, this book could not have been written at all. Their father, George Lewinski, more than once changed his work so that I could pursue mine. Without him, I certainly could never have managed to produce such splendid children and at least a passably good book too. I am deeply grateful.

Andy thinks he won the argument that began in Oxford (this is Andy speaking now), but it's difficult to know. Alison's first child was born just when I was about to tell her the results of a critical test of theory and that trumped my words. In truth, it was the babies who changed our minds more than we've changed each others'. At the University of Washington I have been extremely fortunate to collaborate with Keith Moore, who himself has studied infants since the 1970s. His creative insights have deeply shaped my thinking about cognitive development. Keith and I have spent hours together in front of newborns, nose to nose with the only people who really *can* answer the question of what's innate. It's difficult to match the excitement we've shared in doing psychology with people whose eyes blinked open only a few hours previously. Also at the University of Washington was an extremely alluring speech scientist, Pat Kuhl, who became my wife in about the middle of this protracted conversation. Pat has regularly contributed terrific ideas, and she's kept me honest about perception, or has tried to anyway. Our daughter, Kate, is a magical child. She effortlessly and joyously does what this book struggles to explain. For the guidance of Julian Meltzoff, who first revealed to me the beauty of a good experiment, and Judith Meltzoff, who loved to read and talk, I am grateful.

We didn't need telescopes to study infants, but there were things to buy. The National Institute of Health (HD22514) provided generous support, as did the Center for Human Development and Disabilities and the Virginia Merrill Bloedel Hearing Center. I thank Craig Harris and Calle Fisher, who have helped in innumerable ways, from infant testing to reference checking. Finally, to a generation of scientists who taught us that babies could see, believe, desire, and intend, that they were good minds to consider when considering philosophical matters, I feel profoundly indebted. Without such work by others, this book would not have been written.

1

The Other Socratic Method

Socrates' Problem

About 2,400 years ago Socrates had a problem. The problem was how we could learn about something like virtue from our sensory experience. Socrates' problem is still unsolved. The difficulty is that we typically seem to have highly abstract and complex representations of the world. These range from mathematical to logical to causal knowledge, from our basic understanding of space, time, and objects to our equally basic understanding of our own minds and those of others. No one has ever been able to explain how we could derive these representations from the concrete and simple information of our senses. Rationalists from Socrates himself to Kant to Chomsky resolve this problem by claiming that the abstract structures are there all along. Empiricists from Aristotle to Hume to the connectionists either insist that we can get there from here if we just keep trying or, more skeptically, that there is no there to get to. That is, they end up denying that these abstract and complex representations really exist.

While the significance of Socrates' question has always been recognized, his method of answering it has not been. Socrates' method was to be a developmental psychologist. In Plato's *Meno* Socrates does not just use theoretical or conceptual arguments to bolster his view of knowledge as recollection. Instead, he conducts a developmental experiment. The anonymous slave boy plays as crucial a role in the argument as any of the aristocrats. Socrates poses a series of questions to the boy, questions intended to reveal his underlying knowledge. He asks the boy whether he believes a number of geometrical facts. When the boy answers "yes" to each question, Socrates concludes that the boy has the

abstract representations in question, independent of education and experience. It is this empirical fact that anchors Socrates' arguments.

After 2,400 years of philosophical speculation Piaget tried Socrates' method again. Like Socrates, and for the same philosophical reasons, Piaget wanted to ask children whether they had the same knowledge as adults. But where Socrates' child always said "Yes, O Socrates," Piaget's children always said "Non, M. Piaget." Piaget's empirical work, like Socrates', was an attempt to find out whether children's abstract conceptual structures—their understandings of objects, minds, space, time, and numbers—were indeed the same as those of adults. But where Socrates saw only similarities, Piaget charted consistent differences. And Piaget drew the opposite conclusion. Most of the child's knowledge was radically different from adult knowledge, and that knowledge changed as the child interacted with the world. Therefore the changes in that knowledge must be due to the child's interaction with the world. Piaget attempted to construct a theoretical account of how these changes took place.

Of course, it is difficult to read the *Meno* without a sense that Socrates' questions are not as probing or accurate as we might wish; the "yes"es, we suspect, are imposed as much as they are detected (to be fair, this is, after all, Socrates' general technique with grown-ups too). More recently, Piaget's own ways of asking the questions have also been subject to criticism and revision. Recent empirical work in infancy and early childhood has led us to a very different view of when children say "yes" and when they say "no."

Moreover, this empirical work has led to the rejection of the central tenets of Piaget's theory: cognitive development does not depend on action, there are complex representations at birth, there are no far-reaching domain-general stage changes, young children are not always egocentric, and so on. This is not an indictment of Piaget. Fifty years is an impressive run for any theory, any really good substantive theory ought to be testable, and most really good substantive theories will eventually be overthrown by new empirical work.

One very prevalent reaction to the overthrow of the Piagetian picture has been a return to Socrates' own position. The many early "yes"es seem to indicate that Socrates was right after all, that powerful abstract representations are not derived from experience but are there all along. This reaction is not limited to developmental psychology but rather is

part of the broader contemporary zeitgeist in philosophy and cognitive science. Piaget's theory was, after all, one of the more succesful attempts to explain how concrete sensory experience could lead to the development of abstract complex representations. If Piaget eventually failed to answer Socrates' question, the failure of accounts like classical learning theory or associationism was far more profound.

While this rush toward rationalism has been taking place, an alternative view has gradually been emerging in the developmental literature. The developmental picture is not, after all, simply that we now think children say "yes" when we used to think they said "no." The empirical research also confirms many of Piaget's examples of genuine conceptual change in children and has generated important new instances of such change. Children are still telling us that their knowledge of the world is radically different from our own (Flavell, 1982).

The theoretical position that is emerging to deal with these empirical facts is what we will call "the theory theory." The central idea of this theory is that the processes of cognitive development in children are similar to, indeed perhaps even identical with, the processes of cognitive development in scientists. Scientific theory change is, after all, one of the clearest examples we know of the derivation of genuinely new abstract and complex representations of the world from experience. The model of scientific change might begin to lead to answers to the developmental questions and, more broadly, might begin to answer Socrates' philosophical question.

This theoretical position has been advanced in a number of very different areas of cognitive development, including children's categorizations, their naive understanding of biology and physics, and their understanding of the mind (Carey, 1985, 1988; Karmiloff-Smith, 1988; Karmiloff-Smith & Inhelder, 1974; Gelman & Wellman, 1991; Gopnik & Wellman, 1992, 1994; Gopnik, 1984b, 1988a; Keil, 1989; Perner, 1991; Wellman, 1985, 1990; Wellman & Gelman, 1992). In each of these fields the theory theory has guided experimentation and provided explanations. Often, however, the invocation of the theory theory has seemed like little more than a helpful metaphor. It has not been clear exactly what the theory is supposed to claim or how it could be differentiated from other accounts. Our first aim in this book is to articulate this alternative theoretical position with as much detail and precision as

we can and to show how it can generate specific predictions, predictions that are not made by other theories.

Almost all of the applications of the theory theory have involved older children, and sometimes older children who are explicitly learning scientific concepts. If the theory is supposed to answer Socrates' question, that is, to account for our general capacity to develop new knowledge, it should apply more generally and be true from the beginnings of life. In addition, infancy and early childhood are the periods in which there has been the greatest explosion of new research. So a second aim of this book is to apply the theory theory to explain what we know about infancy and very early childhood.

Augustine's Problem

About 1,700 years ago Augustine had another problem. How did we learn our first words? The philosophical difficulties behind this question parallel Socrates' difficulty. Given the highly arbitrary connections between words and the world, how could we settle so quickly on the right meanings? Like Socrates, Augustine thought the method for answering this question was to be a developmental psychologist. Also like Socrates, his experimental techniques were a bit shaky. He embarked on a retrospective analysis of his own childhood experience. His answer to the question is famous. Augustine reported that adults around him had pointed to objects and said their names and that he had therefore concluded that the names referred to the objects.

After 1,700 years of speculation, serious empirical investigation of this problem began only 20 years ago, with the rise of developmental psycholinguistics. A group of psycholinguists began looking at children's very early words in the mid 1970s (Bloom, 1973; Bowerman, 1978; Nelson, 1973). A landmark in this field was Lois Bloom's book *One word at a time* (1973). Bloom's book, a diary of her child's first words, was significant because it looked at those words with strikingly few theoretical preconceptions. To a remarkable degree Bloom let Allison, her daughter, speak for herself, charting the contexts in which various words appeared and inferring the child's meanings.

Looking at Allison Bloom's first words would have been a shock for Augustine and ought to be similarly shocking for various linguistic and

psychological theories of meaning since then. None of the prevailing linguistic or psychological theories could predict the set of concepts that Allison chose to encode. Allison said things like "allgone," "there," "uh-oh," and "more," and she used these words in ways that were quite different from adult uses. She began using "more," for example, not as a comparative but as an expression of recurrence, a kind of déjà vu. She used the word "allgone" in a dizzying array of contexts: watching a bubble disappear, finishing a bottle of juice, searching for a nonexistent toy cow, encoding generalizations not encoded by any lexical form (including "gone") in the adult language. Allison did use names, as Augustine had predicted, but her uses of those names altered in interesting ways as she grew. At one point in her development she became "name crazy," not simply, as Augustine suggested, associating adult names with words but seeking out names for even the most obscure objects.

Allison Bloom's early words were not simply labels pasted on to convenient objects. Instead, they reflected a universe organized in sharply different ways than the universe of the adults around her. This was a universe in which the temporary disappearance of objects from sight, or the recurrence of an event, was deserving of a word all to itself, out of a linguistic repertoire that included a total of only about 15 words. Allison's early words could not have been learned as Augustine proposed. Allison's language said "no" to Augustine as loudly as the behavior of Piaget's children said "no" to Socrates.

Unlike Piaget's results, however, the divergent results about early language were not accompanied by any clear theoretical alternative. For 20 years the theoretical accounts of early word learning have made assumptions very similar to Augustine's, even if the details of the theory differed. Almost all of the many accounts of early meanings assumed that the child's first words were object names. Moreover, the rush toward rationalism has, of course, been more pervasive in developmental psycholinguistics than even in cognitive development. A mix of Chomsky and Augustine has become the prevailing account of very early language. On this view, even before they have any experience of language, children assume that the first words refer to objects, in the same way that adult names refer to objects (see, for example, Gentner, 1982; Markman, 1989). This is part of a more general view that assumes strong continuities between the semantic structures of adults and children and that

In the second part of the book we will apply the theory theory to the specific case of cognitive and semantic development in infants and very young children. We will sketch a tentative account of the succession of theories and theory changes in three important domains—the understanding of appearances, the understanding of actions, and the understanding of kinds—from birth to about age three. We will focus on the changes in children's problem solving and language that occur at about 18 months. We will argue that these changes can be fruitfully considered to be theory changes, and we will present both linguistic and behavioral evidence to support this view.

In the third part of the book we will consider the implications of the theory theory for broader questions about language and thought. In chapter 7 we will present evidence suggesting that children's early words consistently encode concepts that reflect theory changes and that may be quite different from the concepts encoded in the adult language. Our strongest evidence comes from empirical relationships between specific emerging problem-solving abilities and specific related semantic developments. All this evidence suggests that children's very particular conceptual discoveries play an important role in shaping their early language.

Another line of evidence and argument is relevant to our claim that linguistic input may itself structure conceptual change and discovery. The primary evidence for this claim comes from cross-linguistic studies and studies of individual differences in input. These findings suggest that there is a bidirectional interaction between semantic and conceptual development. Children's early meanings are a joint product of their own cognitive concerns and the cognitive structures already developed by adults.

Finally, in the last chapter we will concentrate on some of the consequences of these views for more general accounts of cognition and meaning in cognitive science and philosophy. We suggest that this evidence supports a version of what is sometimes called semantic holism. More generally, we argue that if the theory theory is correct, we should revise some of the basic asumptions of cognitive science. Cognitive science has focused on an attempt to give a general account of the representations and rules that constitute our adult knowledge of the world. We suggest instead that cognitive science should focus on the dynamic processes by which these rules and representations can be transformed. This chapter too will be quite philosophical.

We have every hope and expectation that this book will leave the reader with the impression that we are simply fascinated by infants and toddlers. When we see small children in the room, even at a dinner party like Meno's, we are unable to resist talking to them. However, the motivation for our research goes beyond this fascination. Ultimately, our reason for watching and talking to children is the same as Socrates'. The most central questions in cognitive science are questions that only they can answer.

I

The Theory Theory

In this book we will argue that children's conceptual structures, like scientists', are theories, that their conceptual development is theory formation and change, and that their semantic development is theory-dependent. In the next two chapters we will elaborate and defend this claim in general terms, and then we will consider specific cases in succeeding chapters. The best argument for any empirical claim is, of course, the data. However, in the case of the theory theory, a number of prima facie objections to the idea might be made. So we will begin by dealing with some of these objections, clarifying just what the theory theory is claiming, and differentiating it from other theories. We first want to show what the idea is, and that the idea is at least plausible, before we try to show that it is true.

2

The Scientist as Child

But Surely It Can't Really Be a Theory?

The claim that children construct theories is often greeted by scientists, philosophers, and psychologists with shocked incredulity. Surely, they cry, you can't really mean that mere children construct theories, not real theories, the kind of theories that we—that is, we serious, grown-up scientists, philosophers, and psychologists—construct with so much sweat and tears. Injured *amour propre* aside, these foes of the theory theory point to a number of differences between children and scientists. Scientists are supposed to be consciously, in fact, self-consciously, reflective about their theory-forming and theory-confirming activities. They talk about them, and they are part of the scientific stream of consciousness. Only a few adult humans become scientists (there is a division of labor), and they only do science part of the time. They do so in a structured institutional setting in which there is much formal interaction with other scientists. Scientific theory change takes place within the scientific community, and a single change may take many years to be completed.

Obviously, none of these things is true of children. For example, 18-month-olds don't talk about the fact that they are formulating or evaluating theories, and they certainly don't publish journal articles, present conference papers, attend seminars, or attempt to torpedo the reputations of those who disagree with them. All children develop theories. Conceptual change in children takes place within a single individual and takes place relatively quickly: children may develop and replace many theories in the space of a few months or years. Insofar as these particular types of phenomenology and sociology are an important part of theory formation and change in science, whatever the children are doing is not science. Given these plain differences, we might ask whether

the idea that children form theories is anything more than a vague metaphor. For the theory theory to be more than just a metaphor, there has to be some interesting, substantive *cognitive* characterization of science, independent of phenomenology and sociology. Is it plausible that science has this kind of cognitive foundation and that it is similar to the cognitive processes we see in children?

We might imagine that we could turn to the philosophy of science for a simple answer to this question. But philosophers of science have really only begun to consider the question themselves. Historically, the philosophy of science has been riven by conflicts between two very different traditions. One tradition has seen philosophy of science as a normative enterprise. Its job is to prescribe ways in which scientists can do things that will lead to the truth. Classically, many philosophers of science identified the actual practice of scientists with this normative project: they assumed that at least most scientists most of the time did what they self-consciously calculated as most likely to lead to the truth. (Philip Kitcher [1993] has recently called this view "Legend.") Moreover, for many philosophers of science in this tradition, the normative project was seen as an essentially logical or mathematical one. Just as formal, deductive logic gave us a way of guaranteeing the truth of certain types of inferences, so we might be able to construct a logic that would guarantee the truth of scientific inferences. (Some of the classic references in this tradition include Hempel, 1965; Nagel, 1961; and Popper, 1965.)

This tradition was notoriously challenged by an alternative view, starting with Thomas Kuhn (1962). Philosophers of science who looked in some historical detail at the actual practices of scientists found a rather different picture than the picture of Legend. The actual practice of science was often characterized by deep divides between proponents of different theories and was highly influenced by apparently accidental sociological facts, such as the professional power of proponents of particular ideas. This view led, in some circles, to the position that there was little relation between the actual practice of science and the normative project of finding the truth. In its most extreme form (e.g., Feyerabend, 1975), this school suggested that sociology was all there was to science, that there was, in fact, no truth to find.

These historical facts led to a standard view in the philosophy of science in which cognitive and psychological factors played little role.

The view is still prevalent. It emphasizes the sociological institutions of science, on the one hand, and on the other hand, the logical structure of explicit, self-conscious scientific reasoning, or rather, of scientific reasoning as normatively reconstructed by philosophers of science.

A Cognitive View of Science

What might an alternative cognitive view of science be like? Science is cognitive almost by definition, insofar as cognition is about how minds arrive at veridical conceptions of the world. In one sense, scientists *must* be using some cognitive abilities to produce new scientific theories and to recognize their truth when they are produced by others. Scientists have the same brains as other human beings, and they use those brains, however assisted by culture, to develop knowledge about the world. Ultimately, the sociology of science must consist of a set of individual decisions by individual humans to produce or accept theories. Scientists converge, however painfully and slowly, on a single set of decisions. The view that is the consequence of these decisions converges on the truth about the world. Scientists must be using human cognitive capacities to do this. What else could they be using?

The assumption of cognitive science is that human beings are endowed by evolution with a wide variety of devices—some quite substantive and domain-specific, others much more general and multipurpose—that enable us to arrive at a roughly veridical view of the world. Usually in cognitive science we think of these devices in terms of representations of the world and rules that operate on those representations. At any given time people have some set of representations and rules that operate on these representations. Over time, there are other cognitive processes that transform both representations and rules. Representations and rules may not have any special phenomenological mark; we may not know that we have them, though sometimes we do. They may be, and often are, deeply influenced by information that comes from other people, and they allow us to communicate with others who have similar representations and rules. Nevertheless, they are not merely conventional, and they could function outside of any social community.

We might think of science in terms of such an abstract system of representations and rules. The question that we would ask, then, is whether there are any generalizations to be made about the kinds of

representations and rules that underlie scientific knowledge and the kinds of processes that transform those representations and rules over time. Is there anything distinctive or special about scientific representations and rules, anything that differentiates them from other possible cognitive structures? Moreover, does the epistemological force of science, its ability to get things right, come from the nature of these representations and rules or from some feature of reflective phenomenology or social institutionalization?

A further question, then, would be whether these representations and rules are similar, or indeed identical, to those we observe in children, and whether changes in these rules and representations are like the changes we see in cognitive development. This might be true even if the phenomenology and social organization of knowledge in children and scientists are quite different. And it might be particularly likely to be true if, in fact, the specific phenomenology and sociology of science are not a necessary condition for its epistemological force.

These seem to us like straightforward and important questions. It might, of course, turn out that there is, in fact, no distinctive or interesting characterization of the representations and rules that underlie scientific knowledge. The formation and confirmation of scientific representations might be completely unconstrained, the result of some whimsical process of "genius" or "insight." It might be that the phenomenology and sociology of science really do the interesting work. It might turn out that there is little relationship between the representations and rules of scientists and those of children. Is it worth trying to find out if there is such a relationship? The detailed empirical work is what we will ultimately have to turn to, but the project is, we think, more plausible and promising than it might seem on the standard view. A cognitive view of science, and in particular, a view that identifies cognitive change in science and childhood, might provide at least a partial explanation of the most important thing about science, namely that it gets things right. In contrast, it is difficult to see how the phenomenological and sociological features of science could explain its epistemological potency.

Recent work in the philosophy of science presents a dilemma. Science is an activity that is performed by human beings in a social context and that proceeds in various and haphazard ways. But it nevertheless manifests a kind of logic and converges on a truthful account of

doing science is logically possible, it would seem sensible to look at how humans do it, since they are the only successful scientific creatures we know of. On the other hand, knowing something about how it could be done is likely to inform our guesses about how evolution actually did it.

Naturalistic Epistemology and Development: An Evolutionary Speculation

The idea that science is related to our ordinary cognition and that both science and ordinary cognition work for evolutionary reasons is not, of course, new. It is the basic idea behind the "naturalistic epistemology" of Quine and others (Quine & Ullian, 1970; Goldman, 1986; Kornblith, 1985). We want to propose, however, a specific version of the naturalistic-epistemology story. This view also might be a reason for supposing that the structures of science are particularly likely to be similar to those involved in cognitive development. On this view, there might actually be a closer link between science and childhood cognition than between science and our usual adult cognitive endeavors.

Let's go back for a minute to the basic idea of cognitive science. We are endowed by evolution with devices for constructing and manipulating rules and representations, and these devices give us a veridical view of the world. Here is an interesting evolutionary puzzle: Where did the particularly powerful and flexible cognitive devices of science come from? After all, we have only been doing science in an organized way for the last 500 years or so; presumably they didn't evolve so that we could do that. We suggest that many of these cognitive devices are involved in the staggering amount of learning that goes on in infancy and childhood. Indeed, we might tell the evolutionary story that these devices evolved to allow human children, in particular, to learn.

A number of writers have recently suggested proposals for an evolutionary account of cognition (Barkow, Cosmides & Tooby, 1992). The view that cognition evolved has been associated with a particular strongly modular and nativist account of cognition. There is, however, no reason to identify the general claim that evolution is responsible for cognitive structure with a modular view. Moreover, the evolutionary arguments for these claims are typically extremely weak. They simply consist of the speculation that a particular trait might have been helpful to an organism in an environment, in fact, in a hypothetical past envi-

as a sort of extended stay in a center for advanced studies, with even better food-delivery systems.

Science as Horticulture

It is an interesting empirical question as to how much of this epistemological activity survives in ordinary adult life. Perhaps not much. Once, as children, we have engaged in the theorizing necessary to specify the features of our world, most of us most of the time may simply go on to the central evolutionary business of feeding and reproducing. But, we suggest, these powerful theorizing abilities continue to allow all of us some of the time and some of us, namely professional scientists, much of the time to continue to discover more and more about the world around us.

On this view, we could think of organized science as a special cultural practice that puts these cognitive capacities to use to solve new kinds of problems, problems that go beyond the fundamental problems we all solve in the first 10 years or so of life. This very fact almost certainly means that scientists will face problems and find solutions that we will not see in childhood. For example, it is characteristic of the child's problems that the evidence necessary to solve them is very easily and widely available, within crawling distance anyway. It is characteristic of scientific problems that the evidence necessary to solve them is rather difficult to obtain. Formal science quite characteristically applies cognitive processes to things that are too big or too small, too rare or too distant, for normal perception to provide rich evidence. Children, in contrast, typically make up theories about objects that are perceptible, middle-sized, common, and close (including, of course, people). This fact about science raises special problems and leads to special solutions. Often these solutions involve particular social institutions.

Moreover, scientists may themselves add to or even revise their theory forming and testing procedures in the light of further experience (though most likely at least older children do the same). For example, evolution may not have given us very good techniques for dealing with probabilistic information (see Kahneman, Slovic & Tversky, 1982) and we may have to invent cognitive prostheses like statistics to do so. Our hypothesis, however, is that the most central parts of the scientific enterprise, the basic apparatus of explanation, prediction, causal attri-

bution, theory formation and testing, and so forth, is not a relatively late cultural invention but is instead a basic part of our evolutionary endowment.

We might think of formal science as a sort of cognitive horticulture. Horticulturalists take basic natural processes of species change—mutation, inheritance, and selection—and put them to work to serve very particular cultural and social ends in a very particular cultural and social setting. In the sixteenth century horticulturalists bred roses to look like sixteenth century women (like the alba rose Cuisse d'une Nymphe Emue), in the mid twentieth century horticulturalists bred roses to look like mid twentieth century cars (like the hybrid tea rose Chrysler Imperial), and in the late twentieth century horticulturalists bred roses to look like pictures of sixteenth-century roses (like the English rose Portia). In one sense, an explanation of the genesis of these flowers will involve extraordinarily complex and contingent cultural facts. But in another sense, the basic facts of mutation, inheritance, and selection are the same in all these cases, and at a deeper level, it is these facts that explain why the flowers have the traits they do.

In the same way, we can think of organized science as taking natural mechanisms of conceptual change, designed to facilitate learning in childhood, and putting them to use in a culturally organized way. To explain scientific theory change, we may need to talk about culture and society, but we will miss something important if we fail to see the link to natural learning mechanisms.

There is an additional point to this metaphor. Clearly, horticulture was for a long time the most vivid and immediate example of species change around. And yet precisely because it was so deeply embedded in cultural and social practices, it seemed irrelevant to the scientific project of explaining the origin of species naturalistically. It was only when Darwin and then Mendel pointed out the underlying similarities between "artificial" and natural species change that these common natural mechanisms became apparent. Similarly, science has been the most vivid and immediate example of conceptual change around (particularly since most philosophers hang out with scientists more than with children). Its cultural and social features have distracted us from looking at it in naturalistic terms. Looking at the similarities between conceptual change in children and conceptual change in science may similarly yield common natural mechanisms.

Objections: Phenomenology

With this cognitive perspective in mind, we can turn back to the real differences between scientists and children. Do these differences undermine the idea that there are deep cognitive similarities between the two groups?

Take the question of phenomenological differences first. Scientists appear to be more consciously reflective about their theorizing than children. But it is difficult to see, on the face of it, why conscious phenomenology of a particular kind would play an essential role in finding things out about the world. A characteristic lesson of the cognitive revolution is that human beings (or for that matter, machines) can perform extremely complex feats of information processing without any phenomenology at all. It is rather characteristic of human cognition that it is largely inaccessible to conscious reflection. There are various speculations we might offer about the role of consciousness in cognition. At the moment, however, we must be more impressed by how little relation there seems to be between cognition and consciousness, rather than by how much. Why should this be different in the case of scientific knowledge?

Moreover, the actual degree of conscious reflection in real science is very unclear. When asked about the stream of consciousness that accompanied his work, Jerry Fodor is alleged to have replied that it mostly said, "Come on, Jerry. That's it, Jerry. You can do it," and this seems reasonably true of much scientific experience. Certainly if we accept, say, Kepler's (1992) writings as a sample of his stream of consciousness, it seems unlikely that we would want to take such a stream as a direct representation of the cognitive processes involved in Kepler's theory construction. Indeed, scientists' own accounts of their theorizing activities are often met with indignant dismay by philosophers of science.

It is true that scientists articulate their beliefs about the world or about their fields of scientific endeavor. So do children, as we will see. But scientists do not typically articulate the processes that generate those beliefs or that lead them to accept them, nor are they very reliable when they do. The reflective processes are really the result of after-the-fact reconstructions by philosophers of science. If it's not very likely that

scientists' phenomenology is a prerequisite for scientific success, it is far less likely that philosophers' phenomenology is.

Of course, scientists may sometimes do philosophy of science. They may, from time to time, be reflective about their own activities and try to work out the structure of the largely unconscious processes that actually lead them to form or accept theories. When scientists engage in this work, they seem to us more like (rather narcissistic) developmental psychologists than children themselves. Moreover, there may be circumstances in which this kind of deliberative self-reflection on their own theorizing practices is a real advantage to scientists, given the particular kinds of problems they face. However, it seems, at least, much too strong to say that such self-reflection is a necessary condition for theory formation and change in science. It seems unlikely that the reflective phenomenology itself is what gives scientists their theorizing capacities or gives the theories their epistemological force.

Finally, it is also not clear that children's phenomenology is radically different from that of adult scientists. The conventional wisdom, or if you prefer, the conventional rhetoric, is to say that children's theories and theory construction must be "implicit" rather than explicit. It is true that young children have a much more limited ability to report their phenomenology than scientists do. But all this means is that we simply don't know very well what their phenomenology is like. We work with very young, barely linguistic infants, and we can't help but be struck by how similar their expressive behavior is to the behavior we normally associate with scientists. Developmentalists are familiar with a characteristic sequence of furrowed brow, intense stare, and bodily stillness, followed by a sudden smile, a delighted glance at the experimenter, and an expression of self-satisfaction verging on smugness as the infant works out the solution. We don't know precisely what sort of internal phenomenology accompanies these expressions, but it seems at least plausible to us that some of what it is like to be an infant with an object-permanence problem is not, after all, so different from what it is like to be a scientist. Certainly, we suspect that the Fodorian stream of consciousness ("What the hell? Damn, this is hard! Hold on a sec. Jeez, I've got it. Boy, am I smart!") is pretty similar in the two cases. In short, there is little indication that particular types of phenomenology, types not shared by scientists and children, are necessary for theory formation or theory change.

particular conventions of deference and trust might meet these particular cognitive problems [Kitcher, 1993]).

It is worth noting that the sociological institutions of science have shifted in the direction of increasing specialization and institutionalization as the problems of science have become more evidentially intractable. The institutional arrangements of Kepler or Newton or even Darwin were very different from those of contemporary scientists, and these earlier scientists were typically much broader in their range of empirical interests. However, it seems difficult to argue that the basic theorizing capacities of current scientists are strikingly superior to those of Kepler or Newton, in spite of the large differences in social organization.

It's easy to see how the division of labor could result from the need for various kinds of evidence and how that structure could lead to particular distinctive problems and patterns of timing in scientific change. What is extremely hard to see is how the hierarchy could lead to the truth or how the division of labor could itself lead to theory formation and confirmation. The division of labor is one consequence of the different problems children and scientists tackle, and maybe it gives scientists an advantage in solving those particular problems. However, this does not imply that the cognitive resources they use to tackle those problems are different.

Moreover, in other respects the child's sociological organization may actually be superior to the scientist's for cognitive purposes. Infants and children have infinite leisure, there are no other demands on their time and energy, they are free to explore the cognitive problems relevant to them almost all the time. They also have a community of adults who, one way or another, are designed to act in ways that further the children's cognitive progress (if only to keep them quiet and occupied). Finally, this community already holds many of the tenets of the theory that the child will converge on and has an interest in passing on information relevant to the theory to the child.

In fact, we might argue that much of the social structure of science is an attempt to replicate the privileged sociological conditions of infancy. Aside from the division of labor, the social hierarchy largely determines who will get the leisure and equipment to do cognitive work and who other scientists should listen to. The infant solves these problems without needing elaborate social arrangements. These are all differences between children and scientists, but again, they do not necessarily

shared Nobel prizes). This convergence to the truth itself is the best reason for thinking that some general cognitive structures are at work in scientific-theory change. Scientists working independently converge on similar accounts at similar times, not because evolutionary theory or the calculus or the structure of DNA (to take some famous examples) are innate, but because similar minds approaching similar problems are presented with similar patterns of evidence. The theory theory proposes that the cognitive processes that lead to this convergence in science are also operating in children.

Objections: Magic

So far we have been focusing on apparent differences between children and scientists and trying to show that they do not invalidate the thesis that common cognitive processes are involved in the two enterprises. We might make a different kind of objection. Consider the following three examples of "explanation."

Francis Bacon is trying to refute Galileo's claim that Jupiter has moons. "There are seven windows given to animals in the domicile of the head, through which the air is admitted to the tabernacle of the body, to enlighten, to warm and to nourish it. What are these parts of the microcosmos: two nostrils, two eyes, two ears and a mouth. So in the heavens, as in a macrocosmos, there are two favourable stars, two unpropitious, two luminaries and Mercury undecided and indifferent. From this and from many other similarities in nature, such as the seven metals, etc., which it were tedious to enumerate, we gather that the number of the planets is necessarily seven" (Warhaft, 1965).

Hapiya, a Zuni Indian is explaining his family history. "One day we were eating when two snakes came towards us right together.... They stand up, start fighting. They was all tangled up, fall down, stand up, tangled up, fall down. We was watching them there; we was interested in watching them. Our grandpa came along on the west side and saw snake tracks. Grandpa got mad, he scold us. 'You should have killed them instead of watching' is what he told us.... 'That's danger, too, someday your family will disappear,' that's what he told us. And it really did, too. About four years later, all my folks disappeared. I was the only one that got left. You know, I got no sister, no brother, nothing. I'm the only one I've got left" (Tedlock, 1992).

Alexei, aged four, is talking to himself, trying to understand where babies come from (the Mommy part is easy). "But what does the Daddy do? He uses his penis. [Takes down his trousers and contemplatively wiggles his own back and forth, observing it attentively.] A penis, ... a pencil, ... a pencil, ... a penis, ... a pencil. [Looks up brightly.] The Daddy uses his penis to draw the baby!"

Faced with these examples of human reasoning, the theory theorist might well be expected to despair. In all three cases we see rather similar kinds of thinking. All three clearly involve deference: central claims in each argument come from authorities rather than experience. In all three there are loose perceptual associations between the premises and conclusion rather than any kind of inference (stars are like parts of the body; the intimate conflict of the snakes is like family conflict; a penis looks [and sounds] like a pencil).

But there is another thing that these examples all have in common, and the last case of Alexei's reasoning brings it out particularly vividly. In all three cases the speakers have no sources of evidence that are relevant to the claims they make. The kind of reasoning Alexei produced is ubiquitous in young children when they answer questions about phenomena of which they are utterly ignorant: Where do babies come from? Why does it get dark at night? Why does the winter come? In fact, these kinds of examples led Piaget to think that young children are intrinsically incapable of logical or causal inference. In similar ways, anthropologists and historians sometimes suggest that people in other places or at other times are fundamentally irrational and certainly very far from possesing the inferential capacities of scientists.

The empirical facts about 4-year-olds, however, show that Piaget was wrong. Suppose instead of asking Alexei about how babies are made, we asked him to explain how his tricycle works or why his friend dragged the stool over to the high cupboard. In these cases he and other 4-year-olds will give a perfectly well-formed causal explanation ("I put my foot on the pedal, and it goes down, and it makes the wheel go around, and the wheel makes it move." "He wanted the cookies, and he thought they were in the cupboard, and so he got on the stool so he could reach the cupboard to get the cookies") (Bullock and Gelman, 1979; Wellman, 1990). (One might also note that Alexei, now 18 years old, did eventually converge on the correct solution.)

These examples reflect an interesting fact about human beings. What happens when the theorizing mechanisms are faced with a causal problem (Why are there 7 moons? Why did one's family all disappear? Where do babies come from?) but have no relevant evidence to operate on? What happens, we suggest, is magic, a combination of narrative, deference, and association. The important thing about these cases is precisely that they involve problems for which the speaker has no evidence: the number of the planets, the death of loved ones, the mystery of conception. In cases where there is evidence, like those of the bicycle and the cupboard, then the genuine theorizing mechanisms can kick in, and we get at least a primitive form of science instead.

There is another interesting fact about these examples. We suggested that science typically applies theorizing processes to cases where relevant evidence is not readily available. It may well be that in just these cases the contrast between prescientific and scientific explanation will be most vivid, and this may misleadingly suggest that scientific methods are fundamentally different from those of ordinary cognition. The explanation a child (or a Zuni or a medieval astrologer) gives of the planet's motions may be wildly different from a scientific explanation. A child's explanation of how a tricycle works, how bread is baked, or why his friend looks in the cupboard for the cookies might be much more similar to a scientific explanation. But in these latter cases scientific explanation would be redundant.

Empirical Advances

The arguments we have advanced so far are really just plausible reasons why cognitive development in childhood might be much like scientific theory change, in spite of the differences between children and scientists. The proof of the theoretical pudding is in the empirical eating. Fortunately, we do not have to provide gastronomic testimonials all by ourselves. Recently, a broader and broader body of empirical evidence in support of this view has accumulated. Cognitive and developmental psychologists have begun to pay more and more attention to the idea that theory change is a model of cognitive development. These empirical projects have been influenced by philosophical ideas, though many of these ideas have come more from other branches of philosophy than from the philosophy of science itself. Philosophers have increasingly

drawn parallels between scientific knowledge and language and our everyday knowledge and language, and psychologists have increasingly found evidence in support of these parallels.

Two lines of development have been particularly important. First, in the philosophical literature Putnam extended arguments from scientific change to our ordinary use of "natural kind" terms, such as "lemon" or "cat." Putnam (1975, 215–271) argued that such terms, rather than picking out specific features or properties of objects, pick out underlying causal "essences" that we conceive as responsible for these features. The notion of "essence" may keep the reference of a term fixed, even when radical conceptual changes, such as those that take place across scientific theory change, may entirely alter the term's intension.

Putnam's argument was based on facts of scientific change, but it was also based on our commonsense intuitions about such facts. In psychological studies of categorization it has become increasingly apparent that Putnam was quite right. Our ordinary categorizations of common objects are best understood in terms of our underlying theories of the objects involved. Adults typically give common names to objects that they think have an underlying common causal nature, rather than those that share superficial perceptual features (Murphy and Medin, 1985). In practice, these decisions are based on adults' commonsense theories of the objects.

If we look at young children's classificatory language and behavior, we see a similar pattern. Even extremely young children appear to organize their categorization in terms of "natural kinds," underlying essences with causal efficacy. Moreover, their decisions about which objects belong to these natural kinds appear to be rooted in naive theories of physics and biology (Carey, 1985; Keil, 1987, 1989; Gelman & Markman, 1986; Gelman & Wellman, 1991).

Most significantly, it is possible to chart qualitative conceptual changes in children's categorizations as their theories are constructed, modified, and revised. This point was first made, and has been made most clearly, extensively, and persuasively, in Carey's seminal work. For example, the child's categorization of an object as an "animal" or as "alive" changes profoundly as the child's "folk biology" changes (Carey, 1985). Perhaps the reason that Putnam's arguments have sometimes failed to impress psychologists is that they depend on intuitions about how reference takes place across theory changes in science. Except, in

really just instances of quite general cognitive structures, such as metaphor schemata, production systems, or connectionist nets. In contrast, the developmentalists have been dissatisfied with such general accounts of cognition and turn to the example of science to argue for much stronger and specific theorylike cognitive structures.

Empirically, then, philosophy of science and cognitive psychology, particularly developmental psychology, have come together by way of rather roundabout routes through philosophy of language and philosophy of mind. However delayed the union, the consummation of a relationship between philosophy of science and cognitive psychology would be a happy event not only from the perspective of cognitive psychology but also from that of philosophy of science itself. From the point of view of cognitive psychology, the example of science gives us a way of dealing with learning, belief formation, and conceptual change. These are perhaps the most thorny unresolved problems in cognitive science. The rest of this book will really be an elaborated defense of the usefulness of the theory theory for cognitive development. From the point of view of philosophy of science, the idea of largely unconscious theorizing devices, designed by evolution for rapid, powerful, and flexible learning, and exploiting logical regularities to that end, might resolve some of the tensions between the more abstract logical characterizations of scientific change and the actual historical evidence.

So the project, as we see it, is not to show that children do science. Instead, we want to argue that the cognitive processes that underlie science are similar to, or indeed identical with, the cognitive processes that underlie much of cognitive development. It is not that children are little scientists but that scientists are big children. Scientific progress is possible because scientists employ cognitive processes that are first seen in very young children.

What Is a Theory?

So far we have been talking about theories in the vaguest of terms, just waving our hands in the direction of the philosophy of science. In fact, the accounts of theories in the cognitive literature have often been rather vague and underspecified. We would like to remedy this by offering a more detailed and precise account. Within the philosophy of science, of course, there is much controversy about what theories are and how to

distinctive features of theories. First, we'll consider the static, structural features of theories, what theories are. Next we'll consider what theories do. Finally, we'll talk about how theories change. (The account we will offer here was developed in collaboration with Henry Wellman, and a version of it may be found in Gopnik & Wellman, 1994.)

Structural Features of Theories

Theories are always constructed with reference to evidence (Nagel, 1961; Lakatos, 1970; Laudan, 1977; Popper, 1965). One of the morals of modern philosophy of science is that there are no strict separations between all theory and all evidence; evidence, it is said, is "theory-laden." Nor is there some foundational level of experience out of which all other kinds of knowledge are built. Still, in any particular example, we can differentiate between a theory and the evidence on which the theory is based. Moreover, the relation between theory and evidence is a distinctive one.

Abstractness

When we say that theories are abstract, we mean that theoretical constructs are typically phrased in a vocabulary that is different from the vocabulary of the evidence that supports the theory. Theories include entities and laws that are postulated or recruited from elsewhere to explain evidence. Gravity is not itself bodies moving in relation to one another; it is a force postulated to explain the behavior of bodies moving in relation to one another. When we postulate a Darwinian species as a theoretical construct (rather than birds, mammals, etc., as empirical types), we define it in terms quite removed from its apparent features. A green stemmed plant and a woody stemmed one may both be ferns because of their reproductive lineage. Kepler's theory of the planets includes elliptical orbits that are notoriously not visible when we look at the stars' motions in the sky. Theories in biology postulated unseen entities with distinctive properties, like viruses and bacteria, to explain visible symptoms of diseases.

Theoretical constructs need not be unobservable. We can, in fact, see bacteria and viruses through a microscope, and the helical structure of DNA can be observed through X-ray crystallography. But they must be appeals to a set of entities removed from, and underlying, the evidential

phenomena themselves. They are entities and laws that explain the data but are not simply restatements of the data.

Coherence

Theoretical constructs do not work independently; they work together in systems with a particular structure. A second characteristic of theories is their coherence. The entities postulated by a theory are closely, "lawfully," interrelated with one another. The classical view of theories captured the coherence of theories by describing two kinds of relations: intratheoretic deductive relations among theoretical entities and "bridge" or "correspondence" relations that connected theories and evidence (Hempel, 1965). More recent conceptions suggest that this is not true; the nature of the theory itself will influence how the theory maps onto the evidence and vice-versa. To say this need not, however, lead us to conclude that theories offer only tautological explanations or redefinitions. There are noncircular ways of specifying both the relations within the theory and the relations between the theory and evidence that allow for an interaction between the two types of laws (Glymour, 1980). This more recent view, in fact, makes the coherent interrelations between parts of the theory even more important.

Causality

A third distinctive feature of theories, related to these two, is their appeal to causality. That is, in theories, we appeal to some underlying causal structure that we think is responsible for the superficial regularities in the data (Cartwright, 1989). Causal relationships are central to theories in two ways. The intratheoretic relations, the laws, are typically interpreted in causal ways. The mass of an object causes other objects to move toward it; selection causes certain mutations to be preserved. But an equally important aspect of theories is that the theoretical entities are seen to be causally responsible for the evidence. The elliptical movements of the planets cause the planets to appear to march across the sky in distinctive ways.

Ontological commitment

Finally, theories make ontological commitments and support counterfactuals (Levi, 1980). An accepted theory is supposed to carve nature at its joints; the theoretical entities and laws are supposed to tell you what

there is and what it must do. As a consequence, theories not only make predictions; they also make counterfactual claims. If the sun and earth were a different distance apart, the earth's orbit would be different; if the moth had evolved in a climate without industrial pollution; it would have a different kind of coloration; and so on. A test of theoreticity, which we will return to later, is the nature of our surprise at violations of the theory. If we are committed to the theory, such violations should strike us not only as surprising but as being impossible and unbelievable in an important and strong way. This differentiates theories from other types of knowledge.

Functional Features of Theories

Prediction

The structural features of theories give them a characteristic sort of predictiveness. A theory, in contrast to a mere empirical generalization, makes predictions about a wide variety of evidence, including evidence that played no role in the theory's initial construction. Kepler's account allows one to predict the behavior of new celestial objects, moons, for example, which were quite unknown at the time the theory was formulated. Theories in biology allow us to predict that antibiotics will inhibit many bacterial infections, including some, like scarlet fever, that present none of the symptoms of an infected wound, or some, like Legionnaires' disease, that were unknown when the theory was formulated. They also allow us to predict that such drugs will be useless against viral infections, even when the symptoms of the viral infection are identical to those of a bacterial one (Nagel, 1961; Hempel, 1965).

Some of these predictions will be correct: they will accurately predict future events described at the evidential level. Others will be incorrect. Since theories go beyond the evidence and are never completely right, some of their predictions will be falsified (Popper, 1965). In still other cases, the theory will make no prediction at all.

In fact, the theory may in some circumstances have less predictive power than would a large set of empirical generalizations. This is because explanatory depth and force do not simply equate with predictive accuracy. We can make predictions about things without explaining them. Celestial navigators and astrologers, for example, noticed certain con-

sistent patterns in the movements of the stars, and made predictions on this basis, without having an explanation for those patterns.

There are two differences between these predictions and those generated by theories. First, a few theoretical entities and laws can lead to a wide variety of unexpected predictions. Second, in the case of a theory, prediction is intimately tied to explanation and causal attribution. The ability to produce wide-ranging predictions is perhaps the most obvious pragmatic benefit of science, and it may also be the most important evolutionary benefit of developing theorizing abilities. The evolutionary value of a system that leads to accurate and wide-ranging predictions should be obvious. In fact, making accurate predictions about the behavior of the world and your fellow organisms is the sine qua non of cognition.

Interpretation

An additional characteristic of theories is that they produce interpretations of evidence, not simply descriptions and typologies of it. Indeed, theories strongly influence which pieces of evidence we consider salient or important (Kuhn, 1977; Lakatos, 1970; Scheffler, 1967). It is notoriously true that theoretical preconceptions may lead a scientist to dismiss some kinds of counterevidence to theoretical claims as simply noise or as the result of methodological failures. This is not necessarily a bad thing. On the contrary, deciding which evidence to ignore is crucial to the effective conduct of a scientific research program. Theory-driven interpretations help to solve what computer scientists call "the frame problem." Theories provide a way of deciding which evidence is relevant to a particular problem. This too might be an evolutionary benefit of theorizing.

Explanation

A third function of theories often mentioned is that they provide explanations (Hempel, 1965; Kitcher & Salmon, 1989). The coherence and abstractness of theories and their causal attributions and ontological commitments together give them an explanatory force lacking in mere typologies of, or generalizations about, the data. Explaining the position of the evening star in terms of Kepler's theory or the properties of plants in terms of their evolutionary history is (at least) cognitively satisfying. Explaining the position of the evening star by saying that it

Dynamic Features of Theories

So far we have been talking mostly about the static features of theories, the features that might distinguish theories from other cognitive structures, such as typologies or schemas. We have only begun to mention theory changes. But in fact the most important thing about theories is what philosophers call their defeasibility. Theories may turn out to be inconsistent with the evidence, and because of this theories change. In fact, a tenet of modern epistemology is that any aspect of a theory, even the most central ones, may change (Quine, 1961; Laudan, 1977). The dynamic features of theories, the processes involved in theory formation and change, are equally characteristic and perhaps even more important from a developmental point of view.

Theories change as a result of a number of different epistemological processes. One particularly critical factor is the accumulation of counterevidence to the theory. Again, Popper's classical views in philosophy of science suggested that this was the defining feature of theory change (Popper, 1965). The theory made a prediction, the prediction was falsified, and the theory was rejected. In fact, the real story is much more complicated. As with the relation between theory and evidence, these complexities have sometimes led to a kind of epistemological nihilism, as if theory change was just a matter of caprice (Feyerabend, 1975). But while a precise specification of theory change may elude us, there are certainly substantive things to be said about how it typically takes place. There are characteristic intermediate processes involved in the transition from one theory to another (Kuhn, 1977; Lakatos, 1970; Laudan, 1977).

The initial reaction of a theory to counterevidence may be a kind of denial. The interpretive mechanisms of the theory may treat the counterevidence as noise, mess, not worth attending to (Lakatos, 1970). At a slightly later stage the theory may develop ad hoc auxiliary hypotheses designed to account specifically for such counterevidence. Auxiliary hypotheses may also be helpful because they phrase the counterevidence in the accepted vocabulary of the earlier theory. But such auxiliary hypotheses often appear, over time, to undermine the theory's coherence, which is one of its strengths. The theory gets ugly and messy instead of being beautiful and simple. The preference for beautiful theories over ugly ones (usually phrased, less poetically, in terms of simplicity criteria) plays an additional major role in theory change.

the flaws of the Ptolemaic accounts and uses the idea of heliocentrism to deal with them (other planets revolve around the sun, which revolves around the earth). But Brahe failed to accept the central idea that the earth itself goes round the sun. Only with Kepler is there a really coherent heliocentric account that deals both with the anomalies and with the earlier data itself. And while, strictly speaking, experimentation on the heavenly bodies was impossible, periods of intense and detailed observation in this transitional period provided a far richer empirical base than had been available before.

Theories in Childhood

We want to claim that infants and young children have cognitive structures like those we have just been describing. All these characteristics of theories ought also to apply to children's early cognitive structures if these structures are really theoretical. That is, children's theories should involve appeal to abstract theoretical entities, with coherent causal relations among them. Their theories should lead to characteristic patterns of predictions, including extensions to new types of evidence and false predictions, not just to more empirically accurate predictions. Their theories should also lead to distinctive interpretations of evidence: a child with one theory should interpret even fundamental facts and experiences differently than a child with a different theory. Finally, their theories should invoke characteristic explanations phrased in terms of these abstract entities and laws. This distinctive pattern of prediction, interpretation, and explanation is among the best indicators of a theoretical structure and the best ways of distinguishing the theory theory from its developmental competitors.

Different aspects of theories will be apparent at different stages of cognitive development. At the very early stage that we are concerned with here, we will emphasize several structural and functional aspects of theories that have correlates in the behavior of infants and young children. These kinds of behavior provide evidence for the theory theory.

Perhaps the most significant piece of evidence is the distinctive pattern of infant predictions. If 18-month-olds develop a theory, we expect them to produce a wide array of new predictions at the same time. In particular, we should see them make predictions even in cases in

which simple empirical generalizations would fail. We should see something like inductive and deductive inferences.

The second type of evidence is the pattern of interpretation. If 18-month-olds are using a theory, we expect them to misinterpret and misuse evidence that contradicts the theory. Even if available evidence might solve some pragmatic problem, children in the grip of a theory might ignore the evidence.

A third type of evidence comes from the extensions of early words. If these words encode theoretical concepts, they should give a unifying characterization to events and objects with quite different superficial perceptual features. In particular, they should pick out groups of objects or events with similar causal structures, and these groupings should depend on, and be linked to, the child's particular theories.

Finally, with these very young preverbal children, direct evidence of explanation is obviously hard to come by. We will see, however, that infants show affective and motivational patterns strikingly like those involved in explanation for adults. In making correct theoretical predictions, infants show a kind of motivation and satisfaction that goes well beyond any immediate functional or social reward. Infants seem to have cognitive orgasms.

We also propose that the dynamic features we have described should be apparent in children's transitions from one theory to a later one. Children should initially ignore certain kinds of counterevidence, then account for such evidence with auxiliary hypotheses, then use the new theoretical idea in limited contexts, and only finally reorganize their knowledge so that the new theoretical entities play a central role. When the new theory is, as it were, under construction, they should engage in extensive experiments relevant to the theory and collect empirical generalizations. Over a given developmental period we should be able to chart the emergence of the new theory from the earlier one, and we should be able to predict a period of some disorganization in between.

Theories as Representations

So far we have been using the same kind of language as philosophers of science to describe theories and theory change, and we will continue to use this language in describing children. We could, however, translate this language into the theoretical parlance of representations and rules

more familiar in cognitive science. A person's theory is a system that assigns representations to inputs just as one's perceptual system assigns representations to visual input or one's syntactic system assigns representations to phonological input. The representations that it assigns are, however, distinctive in many ways, just as perceptual and syntactic representations are distinctive. We can capture these distinctive structural features by talking about the specific abstract, coherent, causal, ontologically committed, counterfactual supporting entities and laws of the theory, just as we talk about phrase structures when we describe syntactic representations (Chomsky, 1980) or $2\frac{1}{2}$-dimensional sketches when we talk about perceptual representations (Marr, 1982). The representations are operated on by rules that lead to new representations; for example, the theory generates predictions. There are also distinctive functional relations between the theoretical representations and the input to them; theories predict, interpret, and explain data.

We know that the input to our perceptual or syntactic systems is provided by our sensory systems. Exactly what is the input to our theory systems? On one view, we might want to propose that other representational systems translate sensory information into some higher level of ordinary, primary, atheoretical knowledge. On this view, not all our representations are assigned by theories. Rather, an earlier level of processing provides the evidential input to theorizing processes. We could describe a level of "evidence" that is not itself theoretical or affected by the theory but serves as the input to the theory. Some of these systems might correspond to Fodor's (1983) "modules." Alternatively, and more in keeping with the philosophical positions that emphasize the "theory-ladenness" of evidence, the system might simply assign theoretical representations to sensory input without a separate level of evidential representation.

By way of illustration, consider a particular observation, say a particular pattern of tracks in a cloud chamber. On any view there will be some very low level atheoretical perceptual processes that will transform the raw sensory input into some more abstract representational form, say a $2\frac{1}{2}$-dimensional sketch. We might believe that there is also a representational system, distinct from the theory system, that further assigns these inputs an atheoretical "ordinary knowledge" representation. For example, it might represent them as "blue tracks in a white jar on a table." This might then be input to the theorizing system.

Alternatively and, we think, more plausibly, the theory system might itself simply assign the input a particular theoretical representation. It might just represent the input as electrons decaying in a particular way. On this view, there is no atheoretical, evidential, "ordinary knowledge" level of representation, at least not once we get past very low level perceptual processing. All representations will be theory-laden. Even apparently "ordinary" kinds of knowledge, like our knowledge that this is a jar and that is a table, or perhaps even that these are two objects and one is on top of the other, will be the result of the application of everyday theories. For various reasons, this strikes us as a more attractive option than the first one. However, which picture is true is an empirical question, and the truth might differ in different cases.

On both views, the theoretical representation assigned to a particular input would then interact in particular rule-governed ways with the other representations of the theory. Does this input match the predictions of the theory? Do some particular theoretical representations cooccur in a way that suggests some causal link between them, a link not specified in the current theory? The fact that certain representations occurred and not others might lead to changes within the theory itself. This could happen even if there were no separate evidential level of representation outside the theory itself.

The most important and distinctive thing about theories is the fact that the very patterns of representation that occur can alter the nature of the representational system itself. They can alter the nature of the relations between inputs and representations. As we get new inputs, and so new representations, the very rules that connect inputs and representations change. Eventually we may end up with a system that has a completely new set of representations and a completely different set of relations between inputs and representations than the system we started out with. As we will see, this differentiates theories from other kinds of representational systems.

Moreover, this system may be dynamic at yet another level. We have been saying that new inputs to the system change the relation between inputs and representations. It is also possible that the very rules that restructure the relations between inputs and representations may change as a result of new input. That is, as we learn more, we also learn new ways to learn. Certainly this seems to be true in science, and it may also be true in development.

the world because of the unique genius of scientists or because some scientists are more powerful than others. But the kind of system we are talking about will certainly suffer from the the same problems of underdetermination that plague the various proposals put forward by philosophers of science in the normative tradition. Given data, the system will arrive at an answer, and given the same data, different instantiations of the system will (eventually) arrive at the same answer. But other answers will still be logically possible, given the same data.

We might say that the space of relations between the input and output will be very much larger than it will be in a modular system, like the visual perception system, but still smaller than the space of logical possibilities. There will be constraints, though very general constraints, on the kinds of relations between inputs and representations that the system will generate. The constraints correspond to the general assumptions that underlie theory formation: that the world has an underlying causal structure, that the structure is most likely to be the simplest one that corresponds to the data, and so on.

The constraints, on our view, come largely from evolution, and at some level this fact is responsible for their veridicality. Presumably, creatures who constructed representations in different ways in childhood, who did not assume underlying causal structure, did not search for the simplest explanation, did not falsify hypotheses when there was counterevidence, and so on, were at an evolutionary disadvantage. In adulthood, such creatures were less good at predicting which berries were members of a poisonous natural kind, which kinds of minerals would make the most seductive body paint, or when their babies were old enough to be left alone without danger (more germane evolutionary tasks than the proverbial dodging of the sabre-toothed tiger). In this sense nature itself guarantees that the system gets to an understanding of nature.

But, of course, evolution is highly contingent, and for all we know, other systems with different sets of constraints might hit on quite different ways of constructing veridical representations. If quantum-mechanical effects translated into selection pressures, perhaps we would have a cognitive system that derived representations from inputs in quite different ways and would be less frustrated in our attempts to understand the quantum universe.

3

Theories, Modules, and Empirical Generalizations

So far we have been making a case for the similarities between scientific knowledge and cognition and also between scientific change and cognitive development. However, it is important to say that not all knowledge is like science and not all development is like scientific change. The analogy to science would be of little interest if it were. Our claim is that quite distinctive and special cognitive processes are responsible both for scientific progress and for particular kinds of development in children. Other kinds of cognition and cognitive development may be quite different. We further claim that theories and theory changes in particular are related to and reflected in early semantic development. In this chapter we will consider other types of knowledge and other processes that could be responsible for developmental change. These provide a contrast case to the theory theory. Moreover, these types of cognition and cognitive development may interact with theory formation in interesting and important ways.

We also intend this chapter to serve a somewhat more ambitious goal. In the wake of the collapse of Piagetian theory, cognitive development has been a bit of a mess, with almost theories, half theories, pseudo-theories, and theory fragments floating about in the sociological ether. In this chapter we will also try to present a sort of vade mecum, a road map to the developmental possibilities. We don't want to contend that the theory theory is a better account, in general, than the other accounts we will describe, or that they are somehow incoherent or implausible. Indeed, we want to argue for a kind of developmental pluralism: there are many quite different mechanisms underlying cognitive developments. Our aim is to argue that theory formation is one among them, an important one.

Moreover, we will eventually want to argue that theory formation, rather than these other mechanisms, accounts for the particular cognitive and semantic phenomena we will discuss later. To do this, we need to consider what sorts of evidence could discriminate between the theory theory and alternative accounts. Many different mechanisms could be responsible for different phenomena, but some particular mechanism will be responsible for each particular phenomenon, and we want to know which one it is. (As in politics, being a pluralist doesn't mean being a wimp.)

One particularly significant contrast, given the zeitgeist in developmental psycholinguistics, will be the contrast with innate modules, constraints, or other related structures. We will suggest that the representations that result from such innate structures may have some of the static features of theories—they may be abstract, be coherent, make causal attributions of a sort, and even allow predictions and interpretations—but they will not have the dynamic features of theories. In particular, they will be indefeasible; they will not be changed or revised in response to evidence. The other important contrast will be with what we will call empirical generalizations: scripts, narratives, connectionist nets, and other cognitive structures quite closely related to immediate experience. Here the contrast runs in the opposite direction. Empirical generalizations, like theories, are defeasible; they may and indeed frequently will be revised. However, they will not have the abstract and coherent quality of theories, nor will they support explanation, prediction, and interpretation in the same way.

Modules

One serious alternative to the theory theory is the idea that cognitive structures are the consequence of innate modules. According to modularity theories, representations of the world are not constructed from evidence in the course of development. Instead, representations are produced by innate structures, modules, or constraints that have been constructed in the course of evolution. These structures may need to be triggered, but once they are triggered, they create mandatory representations of input (Fodor, 1983).[1]

Often the contrast between modularity accounts and the theory theory is phrased in terms of a more general contrast between nativism

and empiricism. But this general contrast does not capture the distinction accurately. While modules are innate, not all innate structures are modular. We have proposed a distinction between two types of nativism: modularity nativism and "starting-state" nativism (Astington & Gopnik, 1991; Meltzoff & Gopnik, 1993; Gopnik & Wellman, 1994). On the starting-state view, the child is innately endowed with a particular set of representations of input and rules operating on those representations. According to this view, such initial structures, while innate, would be defeasible; any part of them could be, and indeed will be, altered by new evidence. We propose that there are innate theories that are later modified and revised. The process of theory change and replacement begins at birth. To continue Neurath's metaphor, innate theories are the boats that push off from the pier. The boat you start out in may have a considerable effect on the boat you end up with, even if no trace of the original remains.

Innate theories might be important in several ways. If children did not have these initial representations, we might expect them to develop later theories in radically different ways, if they developed them at all. Moreover, the fact that the child begins with an initial theorylike structure, which is then revised and restructured in response to evidence, might help solve some underdetermination problems. Such problems have plagued accounts of conceptual change, both in cognitive psychology and in the philosophy of science. Certainly this type of account seems more tractable than one in which theorylike conceptual structures are constructed from scratch from a disorganized flow of experience.

Modularity nativism, on the other hand, implies a much stronger set of claims. In Fodor's analysis, for example, modules are not only innate; they are also encapsulated. On Fodor's view, the representations that are the outcome of modules cannot be overturned by new patterns of evidence. In Chomsky's (1980) theory of syntax acquisition, the innate universal grammar means that only a very limited set of possible grammars will be developed. It constrains the final form of the grammar in the strong sense that grammars that violate it will never be learned by human beings. The idea that certain syntactic structures are indefeasible is at the very core of the idea of constraints in syntax. Similar claims are often advanced in accounts of perceptual systems (Marr, 1982).

The classic examples of modules are the specialized representations and rules of the visual and syntactic systems. Such modules are supposed to

automatically map given perceptual inputs (retinal stimulation or strings of words) onto more abstract set of representations ($2\frac{1}{2}$-dimensional sketches or phrase structures). They automatically mandate certain "inferences" but not others. Outputs from the system may be taken up by other, more central systems, but the relation is asymmetrical. Information from higher systems cannot reshape the representational structure of the module. Once the module has matured, certain representations of the input will result. Other representations simply cannot be formulated, no matter how much evidence supports them.

What kinds of evidence could differentiate between a modularity theory and a theory theory that includes innate theories? Many kinds of evidence that are commonly adduced to support modularity views can't discriminate between these views and the theory theory. In particular, it may be difficult, if not impossible, to distinguish these views by looking at a single static representational system. At least some of the structural and functional features of theories—their abstractness, coherence, and predictive and interpretive force—can also be found in modules.

In particular, both theories and modular representations may involve abstract entities and rules related to sensory input in only very indirect ways. Also like theories, modules allow predictions that go beyond the input. Moreover, they require the mind to represent input in a particular way—a process that may look like interpretation.

In fact, one of the most interesting and important discoveries of cognitive science is that quite automatized, unconscious, indefeasible representational systems can have a very complex internal structure that looks like the complex structure of an inferential system (see Rock's [1983] discussions of the logic of perception for a particularly elegant and perspicuous example of this). The fact that there is some logic in the relations between input and representations is itself not enough to distinguish modular and theoretical structures. Evolution could seize on these relations precisely because they were, at least roughly, the correct ones.[2]

The crucial evidence differentiating the two views lies in the dynamic properties of modules and theories, in how they develop. However, not all the dynamic features of modules and theories will be different. Again, much of the developmental evidence cited to support modularity can't discriminate between modules and theories.

In particular, the fact that there is some knowledge at birth or in very early infancy is compatible with either an innate initial theory or an innate module. The fact that similar representations develop in different children at about the same age also can't discriminate between the two views. The theory theory proposes that there are mechanisms that, given evidence, alter representations in particular ways. If two children start out with the same theory and are given the same pattern of evidence, they will converge on the same theory at roughly the same time.

Theories and Development

If all these kinds of evidence can't discriminate between modules and theories, is there evidence that can discriminate between them? The crucial differences between the modularity theory and the theory theory concern the relation between experience and conceptual structure, between inputs and representations. According to the theory theory, input is evidence. It radically alters the nature of theoretical concepts. Evidence about planetary movements can lead to the transformation of a geocentric conception of the heavens into a heliocentric one; evidence about Galapagan tortoises can lead from Owen to Darwin. Though the relation between the evidence and the change in the theory is, of course, far from simple, the theory theory proposes that there is something about the world that causes the mind to change, and that this fact ultimately grounds the truth of theories.

There is, in principle, a simple experiment that could always discriminate modularity theory and theory theory. Place some children in a universe that is radically different from our own, keep them healthy and sane for a reasonably long period of time, and see what they come up with. If they come up with representations that are an accurate account of our universe, modularity is right. If they come up with representations that are an accurate account of their universe, the theory theory is right. Unfortunately, given the constraints of the federal budget, not to mention the constraints of conscience, this experiment is impossible. In developmental psychology, observation must often do the work of experiment. We can discriminate between modules and theories by observing the interactions of experience and knowledge, of inputs and representations, in development.

compatible with the evidence will actually be constructed. There are some possible theories that will be constructed by human beings, given a particular pattern of evidence, and some that will not. Formally, this may not be profoundly different from the case of a module with a great many parameters differently triggered by evidence. Empirically, however, there is a world of difference between the degrees of freedom that seem to be available in syntactic and perceptual systems and those available in scientific theories as we know them. Theory formation will turn out to involve some set of particular causal principles that get us from patterns of input to patterns of representation. These causal principles must, however, be deeply and radically different from the "parameter setting" principles that have been proposed for modular systems.

Modularity in Peripheral and Central Processing

The canonical examples of modularity are relatively peripheral systems, such as low-level visual and auditory perception and syntax. It may make sense to think of these systems as indeed indefeasible. This is particularly true for syntax, where modularity arguments have been made most strongly. The most distinctive thing about syntax is that it has no reality outside of linguistic behavior itself. There is no syntactic universe independent of us that we develop new and different ideas about. There is just the way we speak. There are no linguistic scientists who discover that language really has unexpected properties not included in any speaker's grammar. If a child incorrectly infers the rules of a language, it is inaccurate to say that he has got it wrong. Rather, he has simply created a new language. In fact, such cases as the development of creoles are often used to support the hypothesis that syntactic structures are innate (Bickerton, 1981). In these cases children are presented with input that is unlike natural language. In particular, they hear a pidgin language that has been created to allow speakers of different natural languages to communicate. The claim in the literature is that the children create a creole, a new natural language like other natural languages, rather than learning the pidgin language they are exposed to, and that this supports the innateness hypothesis.

Chomsky (1980) himself has muddied the waters by describing syntactic structures as our knowledge of a language, rather than as our ability to speak the language. According to Chomsky, universal grammar

is therefore innate knowledge of language. If knowledge of language is innate, we might think, why not knowledge of other things as well? As Chomsky himself points out, whether we want to call syntactic competence "knowledge" or not is unimportant; there's no copyright on the term. What is important is that this type of knowledge is very different from our knowledge of the external world. Chomsky (1992) sometimes talks as if he thinks his model of the acquisition of syntax may be quite widely applicable to areas of psychology that are more genuinely cognitive (such as our knowledge of the physical or psychological world). It is worth pointing out, however, that Chomsky himself has resisted applying the same sorts of theories to semantics that he has applied to syntax.

Some perceptual phenomena are similarly indefeasible, though in a slightly different way. Unlike grammar, perception does refer to things outside itself. But when at least some perceptions are inaccurate, we tend to think that they are supplemented by beliefs, rather than replaced by better perceptions. We can arrange situations in which the perceptual system makes false inferences about the world. When this happens, we are stuck with the result, at least perceptually. No amount of knowledge will make the Müller-Lyer illusion go away.

Fodor (1983), the modern originator of the modularity idea, appropriately distinguishes between peripheral and central knowledge. Fodor advocates modules but contrasts them with another type of knowledge, "central process" knowledge, which includes scientific knowledge. In a way, our view is quite Fodorian. We also think there is a distinction to be drawn between peripheral modules and central processes, and that central processes include both ordinary everyday concepts and scientific concepts. And we think these everyday concepts are just as much (or, we would say, just as little) innate as scientific ones.[3]

The relative success of modularity accounts in some areas of cognitive science has led, understandably enough, to a tendency to extend those accounts to other, more central types of cognition and language. In particular, some cognitive psychologists, psycholinguists, philosophers, and developmentalists, have tried to extend the model of syntax to semantics, including lexical semantics. On this view, the possible range of semantic structures, the things we can think or in any case express linguistically, are themselves sharply constrained and limited in ways reminiscent of constraints on syntactic structures. Partly as a consequence of this and partly with the help of still further innate constraints on the

relations between syntax and semantics, semantic structures are constrained in much the same way as syntactic structures are.

Such accounts date back to the very beginnings of the Chomskyan revolution, of course, with the early rise and later fall of generative semantics. More recently, however, they have been revived, though in very different forms, by writers like Pinker (1989), Jackendoff (1983), Landau and Jackendoff (1993), Talmy (1985), and Lakoff (1987). (Notice that there is an interesting convergence here of East and West Coast cognitive science.)

Moreover, modularity accounts have been proposed in a variety of cases that appear to involve genuinely conceptual and central knowledge of the world. In particular, Spelke et al. (1992) and Atran (1990) have suggested such a model for at least some aspects of our knowledge of physical objects and living things. Leslie (1988) and Fodor (1992) suggest such a model for our knowledge of the mind. There are even recent accounts proposing that our understanding of quite sophisticated aspects of social life, such as obligation and permission, fits this model (Barkow, Cosmides & Tooby, 1992). These accounts mesh with the accounts proposed in semantics. If there are strong constraints on the possible thoughts we can think and beliefs we can hold, there will also be constraints on the possible things we can say. Spelke et al. (1992) describe their account as "neo-Kantian," and this seems like quite an accurate term for this trend in cognitive science in general. Like Kant, these authors propose that certain conceptual structures are innately given and cannot be overturned by evidence.

Both empirically and conceptually, these applications of modularity to semantics and higher-level cognition have considerably less support than modularity accounts of low-level perceptual, motor, and syntactic abilities. There is, moreover, an important respect in which a modular account of semantics, particularly lexical semantics, and high-level cognition will be different from modular accounts of syntax or perception. The representations of syntax and perception are, at least plausibly, the end of the line. We may indeed have relatively fixed syntactic and perceptual representations. We may not be able to overthrow these structures without abandoning perception and syntax (as we do in scientific and formal languages).

In the case of concepts, beliefs. and words, however, such structures cannot be the end of the line. Our concepts and beliefs and the meanings

of our words can and do change all the time, and do so in radical ways in science. Historically, it was the very fact of these radical changes in science that led to the abandonment of the Kantian view in philosophy. If we want a modularity view of conceptual structure to work, we must have, at the very least, some mechanism by which it feeds into a revisable, defeasible conceptual system, like the systems of science. Constraints must eventually be overthrown, biases rejected, conceptual organs reshaped.

But if there is such a mechanism, then the underdetermination arguments used in support of modularity in the first place become much weaker. The standard claim used to support modularity is that certain kinds of knowledge must be innate, since it is difficult to see how such knowledge could be learned. In answer to this claim, we might ask whether children could acquire these kinds of knowledge if they had a learning mechanism as powerful as that of science. Is a particular concept more underdetermined by evidence than scientific theories are? If we think that the cognitive devices of science are powerful enough to allow such learning to take place, we would need some very clear and strong reasons for believing that children do not have such cognitive devices. We would need to draw a sharp and clear line between our everyday cognitive mechanisms and the cognitive mechanisms of science. Otherwise, the underdetermination arguments would not go through. In any particular case it would then be an open empirical question whether a concept was the result of a module or a theory.

Empirical Generalizations: Scripts, Narratives, and Nets

Recently, nativist accounts of cognition have been more prevalent than empiricist ones, particularly in semantics and in discussions of cognition in infancy. However, there are also accounts in cognitive science and in cognitive development that are much more in the empiricist tradition. These accounts explain cognition in terms of the accumulation of particular pieces of information about the world.

We know, in fact, that even in science there are many cases where we know about things without having a theory about them. We may simply have a collection of observations with some regularities among them. Almost all of scientific medicine has this character. We notice a constant (or even inconstant) conjunction between treatment and cure.

Such conjunctions can be powerful enough to be the bases for whole professions and industries (physicians make a lot more money than physicists). In our discussion of theories in the last chapter, we used the philosophy-of-science term "empirical generalizations" to describe this kind of knowledge.

Some theories in cognitive science have proposed that knowledge consists of these sorts of empirical generalizations. "Scripts" are a good example. Scripts were originally proposed by Schank to provide an account of our everyday knowledge (Schank & Abelson, 1977). Scripts are cognitive structures that are supposed to have some predictive or generalizing force, but they are very different from theories. Nelson (1986) has argued that much of the child's early knowledge is organized into "event structures" much like scripts. Similarly, Bruner (1990) suggests that much of our ordinary knowledge is organized in terms of narratives. Narratives, at least on Bruner's view, are another example of a relatively atheoretical type of knowledge, of a kind of empirical generalization. Narratives may sometimes involve "theoretical" notions like causality, but the real constraints in narratives are simply the unities of time and place. As someone once said about the philosophy of history, a narrative is one damn thing after another. It is likely that some of our knowledge of the world has this character. It consists of a set of fairly narrow generalizations about which events typically follow which.

In the area of development, these theories propose that children combine primitive representations of events into more complicated ordered structures. They discover, for example, that a telephone conversation consists of more than just "hi" and "bye," or that something as apparently simple as eating dinner in fact consists of a number of actions with a characteristic order (no dessert until you finish your peas). This process of combination is often quite context-specific, and factors like familiarity and repetition play an important role.

A rather different empiricist account comes from psychologists working with connnectionist models (Bates & Elman, 1993; Clark, 1993; Karmiloff-Smith, 1992) or dynamic systems (Thelen & Smith, 1994). There are two aspects to connectionist modeling. One is that connectionist systems involve a (somewhat) more neurologically realistic kind of computation than classical computational systems. From this perspective, connectionist systems are simply an alternative way of implementing

macho, humorless, courageous, loyal, and melodramatic [the restaurant also features opera]) and about the kinds of restaurants they would be likely to run. We may never have seen a Klingon restaurant, but our folk theory enables us to predict what one would be like.

Similarly, we typically go beyond scripts precisely when we feel motivated to provide an explanatory account of phenomena. If we were to ask, Why do we eat first and pay later at restaurants? simply repeating the script would be an inadequate reply. We suggest that even the unscientific would grope for some folk-psychological or economic theory rather than being content with the script itself. The same is true of interpretive effects. If we go to a restaurant where we pay first and eat later, the data, though surprising, are acceptable: that's how they do it here. If we went to a restaurant in which the food suddenly appeared on the table out of nowhere, like the magic dining room in Jean Cocteau's *La belle et le bête*, we would reinterpret the data. The restaurant must be run by conjurors good at special effects; the food can't *really* appear out of the blue. The point is that when we want deeper explanatory adequacy or wider predictive power, we turn from scripts to theories, even when we are talking about restaurants.

These differences in the static organization and function of empirical generalizations and theories should allow us to distinguish between the two types of cognitive structures in any particular case. Moreover, while both theories and empirical generalizations are defeasible, the characteristic patterns of developmental change may also be different in the two cases. In the case of a theory, we will typically see a pattern in which the child holds to a particular set of predictions and interpretations for some time; the child has a particular theory. Then we may expect a period of disorganization, in which the theory is in crisis. And finally, we should see a new, equally coherent and stable theory emerge. In contrast, in the case of an empirical generalization, the child manifests a more piecemeal, contextually specific pattern of development. Very familiar and frequent pieces of information are learned first, and other pieces of information are gradually added to this store.

Interactions among Theories, Modules, and Empirical Generalizations

So far we have been considering the epistemological relations between theories, modules, and empirical generalizations. How can we tell when

we have one or another? What kinds of evidence can discriminate between them? There is also another question to ask. Assuming that theories, modules, and empirical generalizations all are possible cognitive structures, and assuming, as the developmental pluralists we are, that all three types of structures exist in us, we can try to understand the developmental relationships among them. How do these kinds of knowledge interact with each other in development?

Modules must help to provide the input to theorizing processes. On any view, there are encapsulated atheoretical processes that take us from patterns of stimulation at the retina to patterns of representation in the visual cortex. There may also be some encapsulated atheoretical phonological and syntactic processes that take us from patterns of stimulation at the cochlea to meaningful linguistic strings. Obviously, both what we see and what we hear from other people are important sources of evidence for theory construction. How far up, as it were, do these modular processes go? Where is the border between modules and theories, between the periphery and the center? If you seriously believed in the modularity of all commonsense conceptual knowledge, if you believed that there is what Spelke et al. (1992) call "core knowledge," the answer would be that the border lies at the dividing line between regular people and scientists. If you believe that our commonsense knowledge is at least partly theoretical, however, then this becomes an interesting and important empirical question.

One way of answering it might be by distinguishing modular perceptual processes and central cognitive ones. We might make a principled distinction between perception and cognition by saying that perceptual processes are those that lead to unrevisable representations, representations that are not, in Pylyshyn's (1984) phrase, "cognitively penetrable." These representations are the input to theories, which assign further representations, in this case highly defeasible representations.

It is important to notice that this way of making the distinction between perception and cognition doesn't map onto our phenomenology in any simple way. As we said earlier, a cognitive view of theories is not committed to any particular phenomenology that accompanies them. In fact, a lesson of the cognitive revolution is that phenomenology does not, in general, recapitulate epistemology. On this view, many representations that are theoretical, in the sense of being revisable and defeasible in theorylike ways, might have much of the direct, vivid phenomenology of perception.

In fact, there is good reason to believe that this is true. We know that in cases of expertise where we have a really extensive, worked-out theory that we regularly employ, claims that call on conceptual and even theoretical knowledge (the black king is in danger of check, the electron decayed, the patient has cancer) may be accompanied by very direct perceptionlike phenomenology. Conversely, it is possible that higher-order cognitive representations might be indefeasible and modular and yet not have the phenomenology of perception at all. For example, this is Spelke et al.'s (1992) suggestion with regard to "folk physics," and this is why she denies that the structures of "core knowledge" are perceptual (see also Gopnik, 1993a, 1993b, and discussion in *Cognitive Development* 8, no. 2 [1993]).

In particular, it seems very likely that such low-level visual representations as those of texture and distance are cognitively impenetrable. On the other hand, it seems very likely that such high-level visual representations as the perception of an object's identity are penetrable and often even theoretical. But there is a large intermediate area, for example, the representations involved in perceptual organization, where the answer is unclear. It might take extensive experimental and developmental work to sort out which representations were the result of modules and which were the result of theories (Rock, 1983).

While modular systems can provide input to theorizing systems, in some respects the two types of structures will simply coexist and develop in parallel and independently. Modules provide input to theories, but they are not replaced by theories. Modular representations may contradict theoretical representations and yet coexist with them. Certain perceptual illusions are, of course, the classic example of this. In something like the Müller-Lyer illusion (figure 3.1) our conceptual system overrides the perceptual system. Nevertheless, the perceptual system seems to continue to generate its modularized representations.

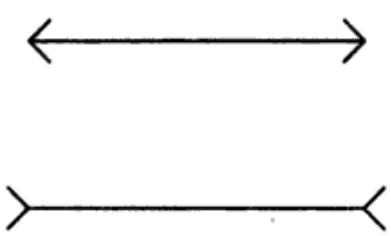

Figure 3.1
The Müller-Lyer illusion. Though the top line looks shorter, because of the outward-pointing arrowheads, the two lines are actually the same length.

Notice, however, that the perceptual illusions generated by modules determine our phenomenology, and perhaps our visual attention or reflexive behaviors, but not our actions or our language. If we want a long stick, we will reach for the stick that we know is longer, not the one that simply looks longer. If someone asks us which stick is longer, we will mention the stick that is really longer. In fact, we need to construct a very special "looks like" vocabulary even to express the perceptual phenomena linguistically at all. In contrast, our conceptual system, our beliefs, our theories are central to both our language and action. When modular and theoretical representations coexist, the theories underwrite a different and much wider range of other mental phenomena than modules.

Sometimes the representations and rules involved in a module may simply remain impenetrable and encapsulated forever. Fourier analysis is a good case in point. We know that at low levels of processing, the visual and auditory system perform fairly complex dynamic analyses of inputs that conform to Fourier analyses (DeValois & DeValois, 1988). We also know that these computational mechanisms are dedicated to these tasks. No process of reflection will give us access to these mechanisms for other purposes. Do we know the principles of Fourier analysis? Whether we want to call the representations of such a module "knowledge" or not is up for grabs; as we said before, there's no copyright on the term. We might say that the eyes, like the heart, have their reasons that reason knows not of. But, however much semantic tolerance we want to extend to modularity theorists, we would at least want to say that the defeasible central representations governing action and language are knowledge if anything is.

There is another possible relation between modules and theories, however. At some stages of development, information from within the originally modular peripheral systems may indeed become available to the central theorizing system. This is not the same as the proposal that the output of the modular systems serves as input to the theorizing systems. Instead, the idea is that at some point the internal structure of the module, its internal representations and rules, become subject to the same kinds of revision and restructuring as more theoretical kinds of representations and rules. In this way a module could be rewritten as an innate theory. We might say that we open up the module, look at what's inside, and turn it into a theory. Karmiloff-Smith (1992) has made extensive and interesting arguments for a role for this kind of "repre-

Words, Thoughts, and Theories

Alison Gopnik and Andrew N. Meltzoff

Words, Thoughts, and Theories articulates and defends the "theory theory" of cognitive and semantic development, the idea that infants and young children, like scientists, learn about the world by forming and revising theories—a view of the origins of knowledge and meaning that has broad implications for cognitive science.

Gopnik and Meltzoff interweave philosophical arguments and empirical data from their own and others' research. The philosophy and the psychology, the arguments and the data, address the same fundamental epistemological question: how do we come to understand the world around us?

Recently the theory theory has led to much interesting research. However, this is the first book to look at the theory in extensive detail and to systematically contrast it with other theories. It is also the first to apply the theory to infancy and early childhood, to use the theory to provide a framework for understanding semantic development, and to demonstrate that language acquisition influences theory change in children.

Alison Gopnik is Professor in the Department of Psychology, University of California at Berkeley. Andrew Meltzoff is Professor in the Department of Psychology, University of Washington.

"Gopnik and Meltzoff review the available evidence in a creative fashion. Anyone looking for an overview of much of the most exciting research in the past ten to fifteen years on early conceptual development would do well to start with their book. Their co-ordination of two lines of evidence that are normally kept separate—children's understanding of the permanence and the appearance of objects—is especially intriguing."
Paul Harris, *Times Literary Supplement*

"Beyond the good science that they contribute to their own idea ..., it is surprising and wonderful how Gopnik and Meltzoff transcend their own field to demonstrate the relevance of their research to other disciplines."
Shaun Gallagher, *Journal of Consciousness Studies*

Learning, Development, and Conceptual Change series
A Bradford Book

Cover art: Jasper Johns, *A Souvenir for Andrew Monk*, 1987. Chalk, charcoal, graphite collage. Collection of Andrew Monk, photograph by Dorothy Zeidman.

The MIT Press
Massachusetts Institute of Technology
Cambridge, Massachusetts 02142
http://mitpress.mit.edu

0-262-57126-9