1

Explaining Explanation

Frank C. Keil and Robert A. Wilson

1.1 The Ubiquity and Uniqueness of Explanation

It is not a particularly hard thing to want or seek explanations. In fact, explanations seem to be a large and natural part of our cognitive lives. Children ask why and how questions very early in development and seem genuinely to want some sort of answer, despite our often being poorly equipped to provide them at the appropriate level of sophistication and detail. We seek and receive explanations in every sphere of our adult lives, whether it be to understand why a friendship has foundered, why a car will not start, or why ice expands when it freezes. Moreover, correctly or incorrectly, most of the time we think we know when we have or have not received a good explanation. There is a sense both that a given, successful explanation satisfies a cognitive need, and that a questionable or dubious explanation does not. There are also compelling intuitions about what make good explanations in terms of their form, that is, a sense of when they are structured correctly.

When a ubiquitous cognitive activity varies so widely, from a preschooler's idle questions to the culmination of decades of scholarly effort, we have to ask whether we really have one and the same phenomenon or different phenomena that are only loosely, perhaps only metaphorically, related. Could the mental acts and processes involved in a three-year-old's quest to know why really be of the same fundamental sort, even if on much smaller scale, as those of an Oxford don? Similarly, could the mental activity involved in understanding why a teenager is rebellious really be the same as that involved in understanding how the Pauli exclusion principle explains the minimal size of black holes? When the domains of understanding range from interpersonal

affairs to subatomic structure, can the same sort of mental process be involved?

Surprisingly, there have been relatively few attempts to link discussions of explanation and cognition across disciplines. Discussion of explanation has remained largely in the province of philosophy and psychology, and our essays here reflect that emphasis. At the same time, they introduce emerging perspectives from computer science, linguistics, and anthropology, even as they make abundantly clear the need to be aware of discussions in the history and philosophy of science, the philosophy of mind and language, the development of concepts in children, conceptual change in adults, and the study of reasoning in human and artificial systems.

The case for a multidisciplinary approach to explanation and cognition is highlighted by considering both questions raised earlier and questions that arise naturally from reflecting on explanation in the wild. To know whether the explanation sought by a three-year-old and by a scientist is the same sort of thing, we need both to characterize the structure and content of explanations in the larger context of what they are explaining (philosophy, anthropology, and linguistics) and to consider the representations and activities involved (psychology and computer science). Even this division of labor across disciplines is artificial: philosophers are often concerned with representational issues, and psychologists, with the structure of the information itself. In addition, disciplinary boundaries lose much of their significance in exploring the relationships between explanation and cognition in part because some of the most innovative discipline-based thinking about these relationships has already transcended those boundaries.

Consider five questions about explanation for which a cognitive science perspective seems particularly apt:

How do explanatory capacities develop?
Are there kinds of explanation?
Do explanations correspond to domains of knowledge?
Why do we seek explanations and what do they accomplish?
How central are causes to explanation?

These are the questions addressed by *Explanation and Cognition*, and it is to them that we turn next.

1.2 How Do Explanatory Capacities Develop?

The ability to provide explanations of any sort does not appear until a child's third year of life, and then only in surprisingly weak and ineffective forms. Ask even a five-year-old how something works, and the most common answer is simply to use the word "because" followed by a repetition or paraphrase of what that thing does. Although three-year-olds can reliably predict how both physical objects and psychological agents will behave, the ability to provide explicit explanations emerges fairly late and relatively slowly (Wellman and Gelman 1998; Crowley and Siegler 1999). But to characterize explanatory insight solely in terms of the ability to provide explanations would be misleading. As adults, we are often able to grasp explanations without being able to provide them for others. We can hear a complex explanation of a particular phenomenon, be convinced we know how it works, and yet be unable to repeat the explanation to another. Moreover, such failures to repeat the explanation do not seem merely to be a result of forgetting the details of the explanation. The same person who is unable to offer an explanation may easily recognize it when presented among a set of closely related ones. In short, the ability to express explanations explicitly is likely to be an excessively stringent criterion for when children develop the cognitive tools to participate in explanatory practices in a meaningful way.

This pattern in adults thus raises the question of when explanatory understanding emerges in the young child. Answering this question turns in part on a more careful explication of what we mean by explanation at any level. Even infants are sensitive to complex causal patterns in the world and how these patterns might be closely linked to certain high-level categories. For example, they seem to know very early on that animate entities move according to certain patterns of contingency and can act on each other at a distance, and that inanimate objects require contact to act on each other. They dishabituate when objects seem to pass through each other, a behavior that is taken as showing a violation of an expectation about how objects should normally behave. These sorts of behaviors in young infants have been taken as evidence for the view that they possess intuitive theories about living and physical entities (e.g., Spelke 1994). Even if this view attributes a richer cognitive structure to infants than is warranted, as some (e.g., Fodor 1998; cf. Wilson and Keil, chap. 4, this volume) have argued, some cognitive structure does cause and explain the

sensitivity. Thus even prelinguistic children have some concepts of animate and physical things through which they understand how and why entities subsumed under those concepts act as they do. We are suggesting that the possession of such intuitive theories, or concepts, indicates at least a rudimentary form of explanatory understanding.

If this suggestion is correct, then it implies that one can have explanatory understanding in the absence of language and of any ability to express one's thoughts in propositional terms. That early explanatory understanding might be nothing more than a grasping of certain contingencies and how these are related to categories of things in turn implies a gulf between such a capacity in infants and its complex manifestation in adults. Certainly, if any sort of explanatory capacity requires an explicit conception of mediating mechanisms and of kinds of agency and causal interactions, we should be much less sure about whether infants have any degree of explanatory insight. But just as the preceding conception of explanation might be too deflationary, we want to suggest that this second view of one's explanatory capacities would be too *inflationary*, since it would seem to be strong enough to preclude much of our everyday explanatory activity from involving such a capacity.

Consider an experimental finding with somewhat older children and with some language-trained apes. An entity, such as a whole apple, is presented, followed by a presentation of the same entity in a transformed state, such as the apple being neatly cut in half. The participant is then shown either a knife or a hammer and is asked which goes with the event. Young children, and some apes, match the appropriate "mechanism" with the depicted event (Premack and Premack 1994; Tomasello and Call 1997). There is some question as to whether they could be doing so merely by associating one familiar object, a knife, with two other familiar object states, whole and cut apples. But a strong possibility remains that these apes and children are succeeding because of a more sophisticated cognitive system that works as well for novel as for familiar tools and objects acted upon (Premack and Premack 1994). If so, is this evidence of explanatory insight, namely, knowing how the apple moved from one state to a new and different one? Mechanism knowledge seems to be involved, but the effect is so simple and concerns the path over time of a single individual. Is this the same sort of process as trying to explain general properties of a kind, such as why ice expands when it freezes?

One possibility about the emergence of explanation is that young children may have a sense of "why" and of the existence of explanations and thereby request them, but are not able to use or generate them much. There is a good deal of propositional baggage in many explanations that may be too difficult for a young child to assimilate fully or use later, but that is at least partially grasped. Perhaps much more basic explanatory schemas are present in preverbal infants and give them some sense of what explanatory insight is. They then ask "why" to gain new insights, but are often poorly equipped to handle the verbal explanations that are offered.

1.3 Are There Kinds of Explanations?

We began with the idea that explanations are common, even ubiquitous, in everyday adult life. A great deal of lay explanation seems to involve telling a causal story of what happened to an individual over time. One might try to explain the onset of the First World War in terms of the assassination of Archduke Ferdinand and the consequent chain of events. There are countless other examples in everyday life. We explain why a friend lost her job in terms of a complex chain of events involving downsizing a company and how these events interacted with her age, ability, and personality, sometimes referring to more general principles governing business life, but often not. We explain why two relatives will not speak to each other in terms of a series of events that led to a blowup and perhaps even explain why it cannot be easily resolved.

Our ease at generating these sorts of narration-based causal explanations, even when they have many steps, contrasts sharply with our difficulty at providing scientific explanations. Explanations in terms of more general laws and principles comprise vastly fewer steps and are cognitively much more challenging. One possible reason may have to do with the closeness between explanations of individual histories and our ability to construct and comprehend narratives more generally, one of the earliest human cognitive faculties to emerge (Neisser 1994; Fivush 1997). By contrast, it is a fairly recent development that people have offered explanations of kinds in terms of principles. Even explanations of various natural phenomena in traditional cultures are often told as narratives of what happened to individuals, such as how the leopard got its spots or why the owl is drab and nocturnal. Are explanations in science therefore of a fundamentally different kind than in normal everyday practice? The answer

is complex, as the essays that follow make clear. It is tempting to think that science does involve the statement of laws, principles, and perhaps mechanisms that cover a system of related phenomena. Yet one must also acknowledge the limits of the deductive nomological model of scientific explanation and the need to conceptualize scientific understanding and practice as something more (or other) than a set of axioms and propositions connected in a deductive pattern of reasoning. In recognizing the limits of the deductive-nomological model of scientific explanation, to what extent do we close the prima facie gap between scientific explanation and the sorts of intuitive explanations seen in young children?

Other sorts of explanations are neither narratives of individual histories nor expositions of general scientific principles. Why, for example, are cars constructed as they are? Principles of physics and mechanics play a role, but so also do the goals of car manufacturers, goals having to do with maximizing profits, planned obsolescence, marketing strategies, and the like. To be sure, these patterns draw on principles in economics, psychology, and other disciplines, but the goals themselves seem to be the central explanatory construct. For another example, we might explain the nature of a class of tools, such as routers, in terms of the goals of their makers. Again such goals interact with physical principles, but it is the goals themselves that provide explanatory coherence. In biology as well, teleological "goals" might be used to explain structure-function relations in an organism without reference to broader principles of biology.

We see here three prima facie distinct kinds of explanation—principle based, narrative based, and goal based—all of which are touched on in the chapters in this book. A key question is what, if anything, all three share. One common thread may involve a pragmatic, coherence constraint that requires that all causal links be of the same sort and not shift radically from level to level. Thus, in a narrative explanation of why Aunt Edna became giddy at Thanksgiving dinner, it will not do to explain how the fermenting of grapes in a region in France caused there to be alcohol in her wine that then caused her altered state. Nor will it do to discuss the neurochemistry of alcohol. It will do to explain the mental states of Edna and those around her that led her to consume large amounts of wine. Similar constraints may be at work in goal-centered and principle-based explanations. We do not yet know how to specify why some set of causal links are appropriate for an explanation and why other equally causal ones are not. We do suggest that common principles may be at work

across all three kinds of explanation; at the least, that question is worth posing and investigating.

1.4 Do Explanation Types Correspond to Domains of Knowledge?

Consider whether there are domains of explanation and what psychological consequences turn on one's view of them. At one extreme, we might think that there are many diverse and distinct domains in which explanations operate. There is a social domain, where our "folk psychological" explanations are at home; there is a physical domain, about which we might have both naive and sophisticated theories; there is a religious domain with its own types of explanatory goals and standards, and so on, with the domains of explanation being largely autonomous from one another. At the other extreme, we might think that these domains are interdependent and not all that diverse. For example, some have proposed that children are endowed with two distinct modes of explanation that shape all other types of explanation they come to accept: an intuitive psychology and an intuitive physical mechanics (Carey 1985). In this view, children's intuitive biology emerges from their intuitive psychology, rather than being one distinct domain of knowledge and explanation among others in early childhood.

It seems plausible that the ability to understand and generate explanations in one domain, such as folk psychology, may have little or nothing in common with the same ability in another domain, such as folk mechanics. The nature of the information to be modeled is different, as are the spatiotemporal patterns governing phenomena in both domains. For example, social interactions have much longer and more variable time lags than do most mechanical ones. While an insult can provoke a response in a few seconds or fester for days, most mechanical events produce "responses" in a matter of milliseconds with little variation across repetitions of the event. At the same time, there may also be overarching commonalities of what constitute good versus bad explanations in both domains and how one discovers an explanation. Again, the essays in this volume explore both dimensions to the issue.

Yet explanations may also be interconnected in ways that call into question the idea that domains of explanation are completely autonomous from one another. Consider how the heart works, a phenomenon whose explanation might be thought to lie within the biological domain. If

pressed hard enough in the right directions, however, the explainer must also refer to physical mechanics, fluid dynamics, thermodynamics, neural net architecture, and even mental states. Explanations might be thought to fall naturally into a relatively small number of domains but, on occasion, leak out of these cognitive vessels. In this view explanations are constrained by domains in that explanations form domain-based clusters, where each cluster is subject to its own particular principles, even if locating the cluster for specific explanations proves difficult or even impossible. Notoriously, the quest for an explanation of sufficient depth can be never ending. "Why" and "how" questions can be chained together recursively; such chains are generated not only by those investigating the fundamental nature of the physical or mental worlds, but also by young children, much to the initial delight (and eventual despair) of parents.

Although, with domains of explanation, we can avoid the conclusion that to know anything we must know everything, we should be wary of thinking of these domains as isolated atoms. To strike a balance between avoiding a need for a theory of everything on the one hand and excessive compartmentalizing, on the other, is one of the key challenges addressed in several of the chapters that follow. The need for such a balance is also related to whether there might be principles that cut across both domains and kinds of explanations, principles that might tell us when a particular causal chain emanating out of a causal cluster has shifted the level or kind of explanation beyond the cluster's normal boundaries and is thus no longer part of *that* explanation.

1.5 Why Do We Seek Explanations and What Do They Accomplish?

What are explanations for? The answer is far more complex and elusive than the question. It might seem intuitively that we seek explanations to make predictions, an answer that receives some backing from the correspondence between explanation and prediction in the deductive-nomological model of explanation and the accompanying hypothetico-deductive model of confirmation in traditional philosophy of science: the observable outcomes predicted and confirmed in the latter are part of the *explanandum* in the former. Yet in many cases, we seem to employ explanations after the fact to make sense of what has already happened. We may not venture to make predictions about what style of clothing will be in

vogue next year but feel more confident explaining why after the fact. If this sort of explanatory behavior occurs with some frequency, as we think it does, a question arises as to the point of such after-the-fact explanations. One possibility, again implicit in many chapters in this volume, is that explanations help us refine interpretative schemata for future encounters, even if prediction is impossible or irrelevant. We may seek explanations from a cricket buff on the nuances of the game, not to make any long range predictions, but merely to be able to understand better in real time what is transpiring on the field and to be able to gather more meaningful information on the next viewing of a cricket match. Here prediction may be largely irrelevant. We may also engage in explanations to reduce cognitive dissonance or otherwise make a set of beliefs more compatible. A close relative dies and, at the eulogy, family members struggle to explain how seemingly disparate pieces of that person fit together. They try to understand, not to predict, but to find a coherent version they can comfortably remember. Simply resolving tensions of internal contradictions or anomalies may be enough motivation for seeking explanations. We suggest here that a plurality of motivations for explanation is needed.

More broadly, we can ask why explanations work, what it is that they achieve or accomplish, given that they are rarely exhaustive or complete. Does a successful explanation narrow down the inductive space, and thus allow us to gather new information in a more efficient fashion? Does it provide us with a means for interpreting new information as it occurs in real time? Given the diversity of explanations, we doubt that there is any single adequate answer to such questions; yet it seems unlikely that a thousand explanatory purposes underlie the full range of explanatory practices. We think that the set of purposes is small and that they may be arrayed in an interdependent fashion. Some explanations might help us actively seek out new information more effectively. Some of those might also help guide induction and prediction. To the extent that we can construct an account that shows the coherence and interrelatedness of explanatory goals and purposes, we can also gain a clearer idea of the unitary nature of explanation itself.

1.6 How Central Are Causes to Explanation?

One final issue concerns the role of the world in general and causation in particular in explanation. At the turn of the century, Charles Sanders

Pierce argued that induction about the natural world could not succeed without "animal instincts for guessing right" (Peirce 1960–1966). Somehow the human mind is able grasp enough about the causal structure of the world to allow us to guess well. We know from the problem of induction, particularly in the form of the so-called new riddle of induction made famous by Nelson Goodman (1955), that the power of brute, enumerative induction is limited. To put the problem in picturesque form, map out any finite number of data points. There will still be an infinite number of ways both to add future data points (the classic problem of induction, from David Hume) as well as connect the existing points (Goodman's new riddle). What might be characterized as a logical problem of how we guess right must have at least a psychological solution because we do guess right, and often.

The idea that we and other species have evolved biases that enable us to grasp aspects of the causal structure of the world seems irresistible. But there is a question as to which of these biases make for explanatory abilities that work or that get at the truth about the world, and how these are related to one another. We might ask whether explanatory devices, of which we are a paradigm, require a sensitivity to real-world causal patterns in order to succeed in the ways they do. Certainly making sense of the world is not sufficient for truth about the world. Both in everyday life and in science, explanations and explanatory frameworks with the greatest survival value over time have turned out to be false. But the sensory and cognitive systems that feed our explanatory abilities are themselves often reliable sources of information about what happens in the world and in what order it happens. Surely our explanatory capacities are doing more than spinning their wheels in the quest to get things right.

While there certainly are explanations in domains where causal relations seem to be nonexistent, such as mathematics or logic, in most other cases there is the strong sense that a causal account is the essence of a good explanation, and we think that this is more than just an illusion. But whether we can specify those domains where causal relations are essential to explanatory understanding, and do so utilizing a unified conception of causation, remain open questions. Philosophers have a tendency to look for grand, unified theories of the phenomena they reflect on, and psychologists often seek out relatively simple mechanisms that underlie complicated, cognitively driven behaviors. Both may need to recognize that the relations between causation and explanation are

complex and multifaceted and may well require an elaborate theory of their own.

Many of the questions we have just raised are some of the most difficult in all of cognitive science, and we surely do not presume that they will be answered in the chapters that follow. We raise them here, however, to make clear just how central explanation is to cognitive science and all its constituent disciplines. In addition, we have tried to sketch out possible directions that some answers might take as ways of thinking about what follows. The chapters in this book attempt, often in bold and innovative ways, to make some inroads on these questions. They explore aspects of these issues from a number of vantage points. From philosophy, we see discussions of what explanations are and how they contrast and relate across different established sciences, as well as other domains. From a more computational perspective, we see discussions of how notions of explanation and cause can be instantiated in a range of possible learning and knowledge systems, and how they can be connected to the causal structure of the world. Finally, from psychology, we see discussions of how adults mentally represent, modify, and use explanations; how children come to acquire them and what sorts of information, if any, humans are naturally predisposed to use in building and discovering explanations. More important, however, all of these chapters show the powerful need to cross traditional disciplinary boundaries to develop satisfactory accounts of explanation. Every chapter draws on work across several disciplines, and in doing so, develops insights not otherwise possible.

The thirteen essays in *Explanation and Cognition* have been arranged into five thematic parts. The chapters of part I, "Cognizing Explanation: Three Gambits," provide three general views of how we ought to develop a cognitive perspective on explanation and issues that arise in doing so. Represented here are an information-processing view that adapts long-standing work to the problem of discovering explanations (Simon); a philosophical view on the psychological differences between science and religion (McCauley); and a view that attempts to connect the perspectives of both philosophers of science and developmental and cognitive psychologists on the nature of explanation (Wilson and Keil).

In his "Discovering Explanations" (chapter 2), Herb Simon views explanation as a form of problem solving. Simon asks how it is that we can discover explanations, an activity at the heart of science, and move

beyond mere descriptions of events to explanations of their structure. He applies his "physical symbol system hypothesis" (PSS hypothesis) to classes of information-processing mechanisms that might discover explanations, and how computational models might inform psychological ones. He also considers patterns in the history and philosophy of science and their relations to structural patterns in the world, such as nearly decomposable systems and their more formal properties, as well as attendant questions about the social distribution and sharing of knowledge.

Robert McCauley explores the relationships between science and religion, and how explanation is related to the naturalness of each, given both the character and content of human cognition as well as the social framework in which it takes place. McCauley's "The Naturalness of Religion and the Unnaturalness of Science" (chapter 3) draws two chief conclusions. First, although scientists and children may be cognitively similar, and thus scientific thought a cognitively natural activity in some respects, there are more significant respects in which the scientific thinking and scientific activity are unnatural. Scientific theories typically challenge existing, unexamined views about the nature of the world, and the forms of thought that are required for a critical assessment of such dominant views mark science as unnatural. Second, an examination of the modes of thought and the resulting products of the practices associated with religion leads one to view religion, by contrast, as natural in the very respects that science is not. Religious thinking and practices make use of deeply embedded cognitive predispositions concerning explanation, such as the tendency to anthropomorphize, to find narrative explanations that are easy to memorize and transmit, and to employ ontological categories that are easy to recognize. These conclusions may help explain the persistence of religion as well as raise concerns about the future pursuit of science.

Our own chapter, "The Shadows and Shallows of Explanation" (chapter 4), attempts to characterize more fully what explanations are and how they might differ from other ways in which we can partially grasp the causal structure of the world. We suggest that traditional discussions of explanation in the philosophy of science give us mere "shadows" of explanation in everyday life, and that one of explanation's surprising features is its relative psychological "shallowness." We further suggest that most common explanations, and probably far more of hands-on science than one might suspect, have a structure that is more implicit and schematic in nature than is suggested by more traditional psychological accounts. We

argue that this schematic and implicit nature is fundamental to explanations of value in most real-world situations, and show how this view is compatible with our ability to tap into causal structures in the world and to engage in explanatory successes. Like Simon, we also consider the importance of the epistemic division of labor that is typically involved in explanatory enterprises.

Part II, "Explaining Cognition," concerns general issues that arise in the explanation of cognition. Its two chapters explore models of explanation used to explain cognitive abilities, locating such models against the background of broader views of the nature of explanation within the philosophy of science. One central issue here is how and to what extent explanation in psychology and cognitive science is distinctive.

Robert Cummins's "'How Does It Work?' versus 'What Are the Laws?': Two Conceptions of Psychological Explanation" (chapter 5), builds on his earlier, influential view that psychological explanation is best conceived not in terms of the Hempelian deductive-nomological model of explanation but rather in terms of capacities via the analytical strategy of decomposition. While the term *law* is sometimes used in psychology, what are referred to as psychological laws are typically effects, robust phenomena to be explained, and as such are *explananda* rather than *explanantia*. Cummins explores the five dominant explanatory paradigms in psychology—the "belief-desire-intention" paradigm, computational symbol processing, connectionism, neuroscience, and the evolutionary paradigm—both to illustrate his general thesis about explanation in psychology and to identify some assumptions of and problems with each paradigm. Two general problems emerge: what he calls the "realization problem" and what he calls the "unification problem," each of which requires the attention of both philosophers and psychologists.

Andy Clark's "Twisted Tales: Causal Complexity and Cognitive Scientific Explanation" (chapter 6) discusses how phenomena in biology and cognitive science often seem to arise from a complex, interconnected network of causal relations that defy simple hierarchical or serial characterizations and that are often connected in recurrent interactive loops with other phenomena. Clark argues that, despite objections to the contrary, models in cognitive science and biology need not reject explanatory schemata involving internal causal factors, such as genes and mental representations. His discussion thereby links questions about the philosophy of science to the practice of cognitive science.

Essays in Part III, "The Representation of Causal Patterns," focus on the centrality of causation and causal patterns within a variety of explanations, continuing a contemporary debate over how causation is represented psychologically. Traditional philosophical views of causation and our knowledge of it, psychological theories of our representation of causal knowledge, and computational and mathematical models of probability and causation intersect here in ways that have only recently begun to be conceptualized.

In "Bayes Nets as Psychological Models" (chapter 7), Clark Glymour focuses on the question of how we learn about causal patterns, a critical component in the emergence of most explanations. Building on developments in computer science that concern conditional probability relations in multilayered causal networks, Glymour considers how a combination of tabulations of probability information and a more active interpretative component allow the construction of causal inferences. More specifically, he argues for the importance of directed graphs as representations of causal knowledge and for their centrality in a psychological account of explanation. This discussion naturally raises the question of how humans might operate with such multilayered causal networks, an area largely unexplored in experimental research. Glymour turns to work by Patricia Cheng on causal and covariation judgments to build links between computational and psychological approaches and to set up a framework for future experiments in psychology.

Woo-kyoung Ahn and Charles Kalish describe and defend a contrasting approach to the study of causal reasoning and causal explanation, what they call the "mechanism approach", in their "The Role of Mechanism Beliefs in Causal Reasoning" (chapter 8). Ahn and Kalish contrast their approach with what they call the "regularity view," as exemplified in the contemporary work of Glymour and Cheng, and stemming ultimately from David Hume's regularity analysis of causation in the eighteenth century. Ahn and Kalish find the two approaches differ principally in their conceptions of how people think about causal relations and in their positions on whether the knowledge of mechanisms per se plays a distinctive role in identifying causes and offering causal explanations. They offer several examples of how mechanistic understanding seems to affect explanatory understanding in ways that go far beyond those arising from the tracking of regularities.

In "Causality in the Mind: Estimating Contextual and Conjunctive Causal Power" (chapter 9), Patricia Cheng provides an overview of her "Power PC theory", where "power" refers to causal powers, and "PC" stands for "probabilistic contrast model" of causal reasoning, an attempt to show the conditions under which one can legitimately infer causation from mere covariation. Cheng employs her theory to suggest that, by instantiating a representation of the corresponding probabilistic relations between covarying events people are able to infer all sorts of cause-and-effect relations in the world. While Glymour (chapter 7) suggests how to extend Cheng's model from simple, direct causal relations to causal chains and other types of causal networks, Cheng herself offers several other extensions, including the case of conjunctive causes.

Paul Thagard's "Explaining Disease: Correlations, Causes, and Mechanisms" (chapter 10) attempts to show that the distance between the two perspectives represented in the the first two chapters of part III may not be as great as the proponents of each view suggest. Thagard focuses on the long-standing problem of how one makes the inference from correlation to causation. He suggests that some sense of mechanism is critical to make such inferences and discusses how certain causal networks can represent such mechanisms and thereby license the inference. His discussion covers psychological work on induction, examines epidemiological approaches to disease causation, explores historical and philosophical analyses of the relations between cause and mechanism, and considers computational problems of inducing over causal networks.

Although several chapters in part I of the book touch on the relationships between cognitive development and science, the two chapters of part IV, "Cognitive Development, Science, and Explanation," explore this topic more systematically. Indeed, the first of these chapters might profitably be read together with McCauley's chapter on science and religion, while the second has links with Wilson and Keil's chapter.

William Brewer, Clark Chinn, and Ala Samarapungavan's "Explanation in Scientists and Children" (chapter 11) asks how explanations might be represented and acquired in children, and how they compare to those in scientists. They propose a general framework of attributes for explanations, attributes that would seem to be the cornerstones of good explanations in science, but that perhaps surprisingly also appear to be the cornerstones of explanation even in quite young children. At the same

time, explanations in science differ from both those in everyday life and from those in the minds of young children, and Brewer, Chinn, and Samarpungavan discuss how and why.

Alison Gopnik addresses the phenomenology of what she calls the "theory formation system," developing an analogy to biological systems that seem to embody both drives and a distinctive phenomenology in her "Explanation as Orgasm and the Drive for Causal Knowledge: The Function, Evolution, and Phenomenology of the Theory Formation System" (chapter 12). In discussing this phenomenology, Gopnik blends together psychological and philosophical issues and illustrates how developmental and learning considerations can be addressed by crossing continuously between these two disciplines. She also brings in considerations of the evolutionary value of explanation, and why it might be best conceived as a drive similar in many respects to the more familiar physiological drives associated with nutrition, hydration, and sex.

In the final part, "Explanatory Influences on Concept Acquisition and Use," two chapters discuss ways in which explanatory constructs influence our daily cognition, either in categorization and concept learning tasks or in conceptual combinations. Explanatory structures seem to strongly guide a variety of everyday cognitive activities, often when these are not being explicitly addressed and when explanations are being neither sought nor generated.

In "Explanatory Knowledge and Conceptual Combination" (chapter 13), Christine Johnson and Frank Keil examine a particularly thorny problem in cognitive science, conceptual combinations. Difficulties with understanding how concepts compose have been considered so extreme as to undermine most current views of concepts (Fodor 1998; cf. Keil and Wilson, in press). Here however, Johnson and Keil argue that framework explanatory schemata that seem to contain many concepts can also help us understand and predict patterns in conceptual combination. The chapter devotes itself to detailed descriptions of a series of experimental studies showing how emergent features in conceptual combinations can be understood as arising out of broader explanatory bases, and how one can do the analysis in the reverse direction, using patterns of conceptual combination to further explore the explanatory frameworks that underlie different domains.

Greg Murphy's "Explanatory Concepts" (chapter 14) examines how explanatory knowledge, in contrast to knowledge of simple facts or other

shallower aspects of understanding, influences a variety of aspects of everyday cognition, most notably the ability to learn new categories. Strikingly, an explanatory schema that helps explain some features in a new category has a kind of penumbra that aids acquisition of other features not causally related to those for which there are explanations. Somehow, explanatory structure confers cognitive benefits in ways that extend beyond features immediately relevant to that structure. Murphy argues that this makes sense, given how often, at least for natural categories, many features are learned that have no immediately apparent causal role. Features that fit into explanatory relations are seen as more typical to a category even when they occur much less often than other explanatorily irrelevant features. Such results strongly indicate that explanation does not just come in at the tail end of concept learning. In many cases, it guides concept learning from the start and in ways that can be quite different from accounts that try to build knowledge out of simple feature frequencies and correlations.

Taken together, these essays provide a unique set of crosscutting views of explanation. Every single essay connects with several others in ways that clearly illustrate how a full account of explanation must cross traditional disciplinary boundaries frequently and readily. We hope that researchers and students working on explanation and cognition in any of the fields this collection draws on will be inspired to pursue the discussion further.

Note

Preparation of this essay was supported by National Institutes of Health grant R01-HD23922 to Frank C. Keil.

References

Carey, S. (1985). *Conceptual chonge in childhood*. Cambridge, MA: MIT Press.

Crowley, K., and Siegler, R. S. (1999). Explanation and generalization in young children's strategy learning. *Child Development*, *70*, 304–316.

Fodor, J. A. (1998). *Concepts: Where cognitive science went wrong*. Oxford: Oxford University Press.

Fivush, R. (1997). Event memory in early childhood. In N. Cowan, ed., *The development of memory*. London: University College London Press.

Goodman, N. (1955). *Fact, fiction and forecast*. Indianapolis: Bobbs-Merrill.

Keil, F. C., and Wilson, R. A. (in press). The concept concept: The wayward path of cognitive science: Review of Fodor's *Concepts: Where cognitive science went wrong. Mind and Language*.

Mandler, J. M. (1998). Representation. In D. Kuhn and R. S. Siegler, eds., *Handbook of Child Psychology.* 5th ed. Vol. 2, *Cognition, perception and language*. New York: Wiley.

Neisser, U. (1994). Self-narratives: True and false. In U. Neisser and R. Fivush, eds., *The remembering Self.* Cambridge: Cambridge University Press.

Peirce, C. S. (1960–1966). *Collected papers*. Cambridge, MA: Harvard University Press.

Premack, D., and Premack, A. (1994). Levels of causal understanding in chimpanzees and children. *Cognition*, *50*, 347–362.

Spelke, E. (1994). Initial knowledge: Six suggestions. *Cognition*, *50*, 431–445.

Tomasello, M., and Call, J. (1997). *Primate cognition*. New York: Oxford University Press.

Wellman, H. M., and Gelman, S. A. (1998). Knowledge acquisition in foundational domains. In D. Kuhn and R. S. Siegler, eds., *Handbook of child psychology.* 5th ed. Vol. 2, *Cognition, perception and language*. New York: Wiley.