Coherence in Thought and Action

Paul Thagard

Coherence in Thought and Action

Paul Thagard

A Bradford Book
MIT Press
Cambridge, Massachusetts
London, England

This book was set Sabon by Best-set Typesetter Ltd., Hong Kong, and was printed and bound in the United States of America.

First printing, 2000

Library of Congress Cataloging-in-Publication Data

Thagard, Paul.
Coherence in thought and action / Paul Thagard.
p. cm.—(Life and mind)
"A Bradford book."
Includes bibliographical references and index.

1. Truth—Coherence theory. I. Title. II. Series.

BD 171 .T45 2000
121—dc21 00-035503

Contents

Preface

This book is an essay on how people make sense of each other and the world they live in. Making sense is the activity of fitting something puzzling into a coherent pattern of mental representations that include concepts, beliefs, goals, and actions. I propose a general theory of coherence as the satisfaction of multiple interacting constraints and show that the theory has numerous psychological and philosophical applications. Much of human cognition can be understood in terms of constraint satisfaction as coherence, and many of the central problems of philosophy can be given coherence-based solutions.

Chapter 1 outlines the importance of the concept of coherence for philosophy and psychology and proposes *cognitive naturalism* as a unified approach to answering philosophical and psychological questions. Chapter 2 develops the cognitive theory of constraint satisfaction as coherence. Chapters 3 and 4 address important philosophical problems concerning the nature of knowledge and reality. Justification of our claims to knowledge is based on five kinds of coherence: explanatory, conceptual, analogical, deductive, and perceptual. These also provide the means to evaluate claims about the nature of reality, for example concerning the existence of the external world, other minds, and God.

Chapter 5 shows the relevance of coherence to philosophical and psychological problems in ethics and politics, arguing that ethical and political judgments are appraisals based on deliberative coherence as well as on the kinds of coherence described in chapter 3. Such appraisals concern not only what to believe, but also what to do, and hence address coherence in action as well as thought. Chapter 6 proposes a new theory of emotional coherence, according to which our appraisals of people, things, and actions emerge from judgments of coherence. It also contends that beauty in science and art is a matter of emotional coherence. Chapter 7 discusses how people who disagree about scientific and other issues can form a consensus on the basis of coherence and communication. Chapter 8 contrasts the coherentist approach to causal inference with probabilistic approaches, particularly Bayesian networks. Finally, chapter 9 suggests some directions for future research on how ideas if about coherence can contribute to progress in philosophy and psychology.

The result, I hope, is a highly coherent theory of coherence. Here briefly is what the book aims to do:

• Provide a far more general and precise account of coherence than has previously been available.

• Increase understanding of how human minds make sense of the way the world is and what to do in it.

• Develop coherence-based answers to central problems in epistemology, metaphysics, ethics, politics, and aesthetics.

• Use ideas about coherence to unify philosophical and psychological problems and to integrate cognition and emotion.

• Understand how consensus can be reached, and identify why it is often difficult to achieve.

• Explain the relation between coherence and probabilistic reasoning.

I hope it all makes sense.

Coherence in Thought and Action

1 *Coherence in Philosophy and Psychology*

At the start of the twentieth century, the disciplines of psychology and philosophy were beginning to separate from each other. Originating in the laboratories of Wilhelm Wundt and William James in the 1870s, experimental psychology had grown rapidly in Germany and the United States. Whereas physics became an experimental subject in the 1600s, it took several more centuries before the investigation of mind also became experimental. The nature and operation of mind had been a central concern of philosophers since Plato, and philosophers should have been excited by the eruption of empirical information. Instead, philosophy went its own way, distancing itself from experimental studies of mind and denying their relevance to traditional problems such as the nature of inference and knowledge.

The two main movements of twentieth century philosophy, analytic philosophy and phenomenology, were explicitly antipsychological. Analytic philosophy became dominant in English-speaking countries, establishing a methodology that emphasized logical or linguistic conceptual analysis as central to philosophical investigation and pushing the study of mind into the background. In Germany and later in France, the philosophical approach of phenomenology, originated by Husserl, set itself the task of describing phenomena of conscious experience in order to grasp their ideal meaning. Both analytic philosophy and phenomenology clearly separate philosophy from empirical psychology, establishing philosophy as a conceptual, nonempirical enterprise.

Although analytic philosophy and phenomenology are still widely practiced and taught, intellectually they have fallen on hard times in recent decades. Both have declined into focusing on internal puzzles and historical retrospectives. In contrast, philosophy of mind and allied areas have been reenergized by regaining contact with empirical psychology, particularly with cognitive psychology, which began to supersede behaviorism in the mid 1950s. Cognitive science has emerged as the interdisciplinary study of mind and intelligence, embracing artificial intelligence, linguistics, anthropology, and neuroscience, as well as psychology and philosophy. Now, at the beginning of a new century, it is clear that psychology and philosophy have many fruitful interconnections.

This book explores one such interconnection involving the role of coherence in thought. I use a computational theory of coherence to illuminate both the psychological task of understanding human thinking and the philosophical task of evaluating how people ought to think. The purpose of this introductory chapter is to explain why coherence is a crucial concept for both philosophy and psychology, and to outline the view I call *cognitive naturalism*, which embraces the symbiosis of psychology and philosophy.

1 COHERENCE IN PSYCHOLOGY

People frequently make inferences about what to believe and what to do. Suppose you are trying to decide whether to buy a used car from someone. You need to be able to infer whether the car is in good condition, partly by relying on your own observations and partly by relying on what the seller says about the car's history, maintenance, and repair records. Whether you believe the seller depends on how trustworthy he or she seems to be, which depends on the inferences you make concerning what kind of person the seller is and whether he or she is telling the truth in this instance. On the traditional account of inference that has been with us since Aristotle, your inferences are a series of steps, each with a conclusion following from a set of premises. Part of your chain of inference might be something like this: The seller looks honest. So the seller is honest. So what the seller says is true. So the car is reliable. So I will buy it.

Another view of inference understands it differently, not as the sort of serial, conscious process just described, but as a largely unconscious process in which many pieces of information are combined in parallel into a coherent whole. On this view, your inference about the car and its seller is the result of mentally balancing many complementary and conflicting pieces of information until they all fit together in a satisfying way. The result is a holistic judgment about the nature of the car, the nature of the seller, and whether to buy the car. Such judgments are the result of integrating the diverse information you have to deal with into a coherent total package. Whether you believe what the seller says about the car will depend in part on what you can infer about the car and vice versa.

Talk of holism and coherence might sound rather mystical, but I am not proposing a kind of New Age cognitive psychology. As chapter 2 describes, coherence-based inference can be characterized just as rigorously as traditional, logic-based inference. Moreover, much of human thinking is naturally understood as coherence-based, in domains as diverse as social impression formation, scientific-theory choice, discourse comprehension, visual perception, and decision making. Later chapters will show how these and other kinds of human thinking can be understood in terms of coherence processes. A precise and psychologically plausible theory of coherence has much to contribute to cognitive, social, and developmental psychology. One benefit, described in chapter 6, is a unified account of cognition and emotion.

2 COHERENCE IN PHILOSOPHY

Philosophy differs from psychology in that it is traditionally concerned with normative questions about how people should think, not just descriptive questions about how they do think. At the center of this normative concern is justification: are we justified in having the beliefs that we have acquired, and how can we justify the acquisition of new beliefs? For many philosophers, justification is a matter of finding the right foundation consisting of a set of indubitable beliefs from which other beliefs can be inferred. Two sources of certainty have been pursued: reason and sense experience. Rationalists such as Plato and Descartes attempted to use reason alone to achieve foundations of knowledge that could provide sources of justification for other beliefs. In contrast, empiricists such as Locke, Berkeley, and Hume took sense experience as the foundation for all knowledge.

Today, it is generally recognized that both of these foundational approaches to justification are failures. There are no indubitable truths of reason of the sort that Plato and Descartes sought, and even if there were, they would be too trivial to provide a basis for all the other things we think we know. Similarly, there are no indubitable truths of sense experience, and sense experience alone is too achieved, but it provides no insight on how to achieve it.

Chapter 2 provides a much more precise account of coherence as constraint satisfaction, along with algorithms for computing coherence. Later chapters show how different kinds of coherence, employing different kinds of representations and constraints, cover the most important areas of philosophical thought. My aim, however, is not just to describe the logic of coherence, but to give a psychologically plausible account of how coherence mechanisms operate in the human mind. A computational and naturalistic account of coherence can help not only with traditional philosophical problems of justification, but also with psychological concerns about how the mind works. Before undertaking that task, however, some preliminary remarks about the relation of philosophy and psychology are in order.

3 WHY PHILOSOPHY ABANDONED PSYCHOLOGY

It is commonly believed that in the nineteenth century psychology emerged from philosophy, just as physics, chemistry, and biology had earlier used experimental methods to develop beyond philosophical speculation. In contrast, Reed (1997) argues that it is more accurate to say that philosophy emerged from psychology. The history of philosophy before 1900 is dominated by figures who approached epistemological and metaphysical issues in tandem with questions concerning the nature of mind: Plato, Aristotle, Descartes, Hobbes, Locke, Berkeley, Hume, Kant, and Mill, to name just a few. For these thinkers, philosophy and psychology clearly were not separate disciplines. Similarly for the founders of experimenof experimental psychology, but there are deeper, more conceptual explanations of why philosophy became antipsychological.

For both Frege and Husserl, avoiding psychology was essential for establishing objective truths. Frege's *Basic Laws of Arithmetic*, published in 1893, began with a diatribe against what he called the "psychological logicians," whom he accused of writing logic books that are "bloated with unhealthy psychological fat that conceals all more delicate forms" (Frege 1964, 24). On his view, "the laws of truth are not psychological laws: they are boundary stones set in an eternal foundation, which our thought can overflow, but never displace" (1964, 13). Knowledge of arithmetic has nothing to do with psychology, Frege claimed, but is purely a matter of logic. Similarly, Husserl in 1913 made a sharp distinction between psychology and his enterprise of "pure phenomenology," which he intended to establish "not as a science of facts, but as a science of essential Being," leading the way to "Absolute Knowledge" (Husserl 1962, 40–41). Logical and phenomenological approaches both promised to provide philosophy with a priori knowledge, which no work tainted with empirical psychology could achieve.

The decades have not been kind to either of these ambitious enterprises. Gödel showed in 1931 that logic was insufficient for the foundations of arithmetic, and indubitable a priori truths of the sort sought by Frege, Husserl, and many other philosophers have been elusive. At best, the only defensible a priori truths are trivialities such as "Not every statement is both true and false" (Putnam 1983). The search for solid foundations for knowledge has undoubtedly failed, and this failure has cast some philosophers into the desperate postmodern conclusion that philosophy is dead and that nothing survives but discourse about discourse. Such despair is unwarranted if one adopts a perspective that is coherentist and naturalistic.

Analytic philosophy and phenomenology attracted followers not only because they offered certainty, but also because they offered methods for making philosophical progress. Logical analysis and phenomenological reflection gave philosophers ways of pursuing foundational goals that sharply demarcated their methods from those of empirical psychologists. Along the way, acute philosophers in both traditions often made interesting and important observations about language, meaning, and life in general, although the results of the core methods of logical analysis and phenomenological reduction have been meager. In recent decades, however, naturalistic approaches have undergone a dramatic revival.

4 COGNITIVE NATURALISM

Naturalistic approaches to philosophy that tie it closely to empirical science are as old as philosophy itself. Precursors of contemporary naturalism include Thales, Aristotle, Bacon, Hume, Mill, Peirce, and countless others. Philosophical naturalists see philosophy as continuous with science in both subject matter and method, rejecting supernatural entities. Naturalism need not, however, reduce philosophy to empirical science, which is highly relevant to normative issues in logic, ethics, and aesthetics but does not fully suffice to settle those issues (see chapter 5).

What distinguishes the movement I call *cognitive naturalism* is its close ties with cognitive science, an interdisciplinary amalgam of psychology, artificial intelligence, neuroscience, and linguistics that originated in the mid 1950s (Gardner 1985). The central hypothesis of cognitive science is that thought can be understood in terms of computational procedures on mental representations. This hypothesis has had enormous empirical success, providing explanations of numerous phenomena of human problem solving, learning, and language use. Although there is considerable dispute within cognitive science concerning what kinds of procedures and representations are most important for understanding mental phenomena, the computational/representational approach is common to current work on how mind can be understood in terms of rules, concepts, analogies, images, and neural networks (see Thagard 1996 for a concise survey).

Mirroring the diversity of approaches to cognitive science, philosophers within the cognitive-naturalist movement draw on different aspects of contemporary psychology, linguistics, artificial intelligence, and neuroscience. But the differences should not obscure the commonalities among philosophers who agree that many traditional philosophical problems are intimately tied with results in the cognitive sciences that have implications for issues in epistemology, metaphysics, and ethics (see, for example, P. S. Churchland 1986; P. M. Churchland 1995; Giere 1988; Goldman 1986; Harman 1986; May, Friedman, and Clark 1996).

Cognitive naturalism contrasts with philosophical approaches that predate the rise of the computational/representational view of mind. Quine is an influential twentieth-century naturalist whose epistemological views display the impact of behaviorist psychology, seen especially in his concern with observable stimuli. Quine's major work, *Word and Object*, was published in 1960 and was strongly influenced by his association with his behaviorist colleague B. F. Skinner, but it ignored the emerging approach of George Miller and Jerome Bruner, who were also at Harvard and who started the Center for Cognitive

Studies in 1960. Quine's naturalistic epistemology is a behaviorist naturalism rather than a cognitive naturalism.

Another naturalistic movement in the twentieth century was the "scientific philosophy" of the logical positivists. However, its leaders, such as Carnap and Reichenbach, followed Frege in rejecting the relevance of empirical psychology to epistemological issues and in basing their theories on formal logic. If human thinking employed the apparatus of symbolic logic, then there would be little difference between logical naturalism and cognitive naturalism. But there is abundant evidence that thought requires mental representations such as concepts and images, and computational procedures such as spreading activation and pattern matching, that go beyond the kinds of structures and inference allowed in the logical framework. Frege would have said, so much the worse for psychology, but the failure of the logicist approach to epistemology does not permit such arrogance.

A third kind of naturalistic epistemology is found in the writings of sociologists such as Latour and Woolgar (1986), who claim to explain the development of science exclusively in terms of social relations such as power. Social naturalism, however, is compatible with cognitive naturalism if it more reasonably offers social explanations as complementary to cognitive explanations of science rather than as alternatives. Examples of how cognitive and social naturalism can be combined can be found in Goldman's (1992) discussion of epistemic standards for social practices, Bloor's (1992) acceptance of a cognitive background to social relations, and my own discussion of cognitive and social explanation schemas for scientific change (Thagard 1999).

Unlike the monolithic social naturalism of Latour and Woolgar, cognitive naturalism is nonexclusionary.

Cognitive naturalism is the rising approach to philosophy that finds close ties between philosophy and the cognitive sciences, including psychology, neuroscience, linguistics, and artificial intelligence. A computational approach to coherence has the potential to provide both a powerful theory of important cognitive mechanisms and a non-foundational solution to philosophical problems about justification.

2 *Coherence as Constraint Satisfaction*

As chapter 1 described, the concept of coherence has been important in many areas of philosophy and psychology. But what is coherence? Given a large number of elements (propositions, concepts, or whatever) that are coherent or incoherent with each other in various ways, how can we accept some of these elements and reject others in a way that maximizes coherence? How can coherence be computed? Answers to these questions are important not only for philosophical understanding and the development of machine intelligence, but also for developing a cognitive theory of the role of coherence in human thinking.

Section 1 of this chapter offers a simple characterization of coherence problems that is general enough to apply to a wide range of current philosophical and psychological applications summarized in section 2. Maximizing coherence is a matter of maximizing satisfaction of a set of positive and negative constraints. Section 3 describes five algorithms for computing coherence, including a connectionist method from which my characterization of coherence was abstracted. Coherence problems are inherently intractable computationally, in the sense that, under widely held assumptions of computational complexity theory, there are no efficient (polynomial-time) procedures for solving them. There exist, however, several effective approximation algorithms for maximizing-coherence problems, including one using connectionist (neural network) techniques. Different algorithms yield different methods for measuring coherence, and this is discussed in section 4.

This chapter presents a characterization of coherence that is as mathematically precise as the tools of deductive logic and probability theory more commonly used in philosophy. The psychological contribution of this chapter is that it provides an abstract formal characterization that unifies numerous psychological theories with a mathematical framework that encompasses constraint-satisfaction theories of hypothesis evaluation, analogical mapping, discourse comprehension, impression formation, and so on. Previously these theories shared an informal characterization of cognition as parallel constraint satisfaction, along with the use of connectionist algorithms to perform constraint satisfaction. The new precise account of coherence makes clear what these theories have in common besides connectionist implementations. Moreover, the mathematical characterization generates results of considerable computational interest, including a proof that the coherence problem is NP-hard (nondeterministic-polynomial-hard) and the development of algorithms that provide nonconnectionist means of computing coherence.

1 CONSTRAINT SATISFACTION

When we make sense of a text, picture, person, or event, we need to construct an interpretation that fits with the available information better than alternative interpretations. The best interpretation is one that provides the most coherent account of what we want to understand, considering both pieces of information that fit with each other and pieces of information that do not fit with each other.

For example, when we meet unusual people, we may consider different combinations of concepts and hypotheses that fit together to make sense of their behavior.

Coherence can be understood in terms of maximal satisfaction of multiple constraints in a manner informally summarized as follows:

- The elements are representations, such as concepts, propositions, parts of images, goals, actions, and so on.
- The elements can cohere (fit together) or incohere (resist fitting together). Coherence relations include explanation, deduction, facilitation, association, and so on. Incoherence relations include inconsistency, incompatibility, and negative association.
- If two elements cohere, there is a positive constraint between them. If two elements incohere, there is a negative constraint between them.
- The elements are to be divided into ones that are accepted and ones that are rejected.
- A positive constraint between two elements can be satisfied either by accepting both elements or by rejecting both elements.
- A negative constraint between two elements can be satisfied only by accepting one element and rejecting the other.
- The coherence problem consists of dividing a set of elements into accepted and rejected sets in a way that satisfies the most constraints.

Examples of coherence problems are given in section 2.

More precisely, consider a set E of elements, which may be propositions or other representations. Two members of E, e_1 and e_2, may cohere with each other because of some relation between them, or they may resist cohering with each other because of some other relation.

We need to understand how to make E into as coherent a whole as possible by taking into account the coherence and incoherence relations that hold between pairs of members of E. To do this, we partition E into two disjoint subsets, A and R, where A contains the accepted elements of E, and R contains the rejected elements of E. We want to perform this partition in a way that takes into account the local coherence and incoherence relations. For example, if E is a set of propositions and e_1 explains e_2, we want to ensure that if e_1 is accepted into A, then so is e_2. On the other hand, if e_1 is inconsistent with e_3, we want to ensure that if e_1 is accepted into A, then e_3 is rejected and put into R. The relations of explanation and inconsistency provide constraints on how we decide what can be accepted and rejected.

More formally, we can define a *coherence problem* as follows. Let E be a finite set of elements $\{e_i\}$ and C be a set of constraints on E understood as a set $\{(e_i, e_j)\}$ of pairs of elements of E. C divides into $C+$, the positive constraints on E, and $C-$, the negative constraints on E. With each constraint is associated a number w, which is the weight (strength) of the constraint. The problem is to partition E into two sets, A and R, in a way that maximizes compliance with the following two *coherence conditions*:

- If (e_i, e_j) is in $C+$, then e_i is in A if and only if e_j is in A.
- If (e_i, e_j) is in $C-$, then e_i is in A if and only if e_j is in R.

Let W be the weight of the partition, that is, the sum of the weights of the satisfied constraints. The coherence problem is then to partition E into A and R in a way that maximizes W. Because *a coheres with b* is a symmetric relation, the order of the elements in the constraints does not matter.

Intuitively, if two elements are positively constrained, we want them either to be both accepted or both rejected.

acceptance, but the connectionist algorithm discussed below in section 4 indicates how such degrees can be computed. To show that a given problem is a coherence problem in the sense of this chapter, it is necessary to specify the elements and constraints, provide an interpretation of acceptance and rejection, and show that solutions to the given problem do in fact involve satisfaction of the specified constraints.

2 COHERENCE PROBLEMS

In coherence theories of truth, the elements are propositions, and accepted propositions are interpreted as true, while rejected propositions are interpreted as false. Advocates of coherence theories of truth have often been vague about the constraints, but entailment is one relation that furnishes a positive constraint and inconsistency is a relation that furnishes a negative constraint (Blanshard 1939). Whereas coherence theories of justification interpret "accepted" as "judged to be true," coherence theories of truth interpret "accepted" as "true." A coherence theory of truth may require that the second coherence condition be made more rigid, since two inconsistent propositions can never both be true, but chapter 4 argues against such a theory.

Epistemic justification is naturally described as a coherence problem as specified above. Here the elements in *E* are propositions, and the positive constraints can be a variety of relations among propositions, including entailment and also more complex relations such as explanation. The negative constraints can include inconsistency, but also weaker constraints such as competition. Some propositions are to be accepted as justified, while others rejected.

considered moral beliefs. Swanton (1992) proposed a coherence theory of freedom based on reflective equilibrium considerations. As in Goodman's view of logical justification, the acceptance of ethical principles and ethical judgments depends on their coherence with each other. Coherence theories of law have also been proposed, holding the law to be the set of principles that makes the most coherent sense of court decisions and legislative and regulatory acts (Raz 1992).

Thagard and Millgram (1995, Millgram and Thagard 1996) have argued that practical reasoning also involves coherence judgments about how to fit together various possible actions and goals. On this account, the elements are actions and goals, the positive constraints are based on facilitation relations (action *a* facilitates goal *g*), and the negative constraints are based on incompatibility relations (you cannot go to Paris and London at the same time). Deciding what to do is based on inference to the most coherent plan, where coherence involves evaluating goals as well as deciding what to do. Hurley (1989) has also advocated a coherence account of practical reasoning, as well as ethical and legal reasoning.

In psychology, various perceptual processes such as stereoscopic vision and interpreting ambiguous figures are naturally interpreted in terms of coherence and constraint satisfaction (Marr and Poggio 1976, Feldman 1981). Here the elements are hypotheses about what is being seen, and positive constraints concern various ways in which images can be put together. Negative constraints concern incompatible ways of combining images, for example, seeing the same part of an object as both its front and its back. Word perception can be viewed as a coherence problem in which hypotheses about how letters form words can be evaluated against each other on the basis of constraints on the shapes and interrelations of letters (McClelland and Rumelhart

1981). Kintsch (1988) described discourse comprehension as a problem of simultaneously assigning complementary meanings to different words in a way that forms a coherent whole. For example, the sentence "The pen is in the bank" can mean that the writing implement is in the financial institution, but in a different context it can mean that the animal containment is in the side of the river. In this coherence problem, the elements are different meanings of words, and the positive constraints are given by meaning connections between words like "bank" and "river." Other discussions of natural-language processing in terms of parallel constraint satisfaction include St. John and McClelland 1992 and MacDonald, Pearlmutter, and Seidenberg 1994. Analogical mapping can also be viewed as a coherence problem. Here two analogs are put into correspondence with each other on the basis of various constraints such as similarity, structure, and purpose (Holyoak and Thagard 1989, 1995).

Coherence theories are also important in recent work in social psychology. Read and Marcus-Newhall (1993) have experimental results concerning interpersonal relations that they interpret in terms of explanatory coherence. Shultz and Lepper (1996) have reinterpreted old experiments about cognitive dissonance in terms of parallel constraint satisfaction. The elements in their coherence problem are beliefs and attitudes, and dissonance reduction is a matter of satisfying various positive and negative constraints. Kunda and Thagard (1996) have shown how impression formation, in which people make judgments about other people based on information about stereotypes, traits, and behaviors, can also be viewed as a kind of coherence problem. The elements in impression formation are the various characteristics that can be applied to people; the positive constraints come from correlations among the characteristics; and the negative constraints come from negative correlations. For example, if you are told that someone is a Mafia nun, you have to reconcile the incompatible expectations that she is moral (nun) and immoral (Mafia). Thagard and Kunda (1998) argue that understanding other people involves a combination of conceptual, explanatory, and analogical coherence.

Important political and economic problems can also be reconceived in terms of parallel constraint satisfaction. Arrow (1963) showed that standard assumptions used in economic models of social welfare are jointly inconsistent. Gerry Mackie (personal communication) has suggested that deliberative democracy should not be thought of in terms of the idealization of complete consensus, but in terms of a group process of satisfying numerous positive and negative constraints. Details remain to be worked out, but democratic political decision appears to be a matter of both explanatory and deliberative coherence. Explanatory coherence is required for judgments of fact that are relevant to decisions, and multiagent deliberative coherence is required for choosing what is optimal for the group as a whole. See the end of chapter 5 for further discussion of coherence in politics.

Table 2.1 summarizes the various coherence problems that have been described in this section. Although much of human thinking can be described in terms of coherence, I do not mean to suggest that cognition is one big coherence problem. For example, the formation of elements such as propositions and concepts and the construction of constraint relations between elements depend on processes to which coherence is only indirectly relevant. Similarly, serial step-by-step problem solving such as finding a route to get from Waterloo to Toronto is not best understood as a coherence problem, unlike choosing between alternative routes that have been previously identified. The claim that much of human inference is a matter of coherence in the

Table 2.1
Kinds of coherence problems

	Elements	Positive constraints	Negative constraints	Accepted as
Truth	Propositions	Entailment, etc.	Inconsistency	True
Epistemic justification	Propositions	Entailment, explanation, etc.	Inconsistency, competition	Known
Mathematics	Axioms, theorems	Deduction	Inconsistency	Known
Logical justification	Principles, practices	Justification	Inconsistency	Justified
Ethical justification	Principles, judgments	Justification	Inconsistency	Justified
Legal justification	Principles, court decisions	Justification	Inconsistency	Justified
Practical reasoning	Actions, goals	Facilitation	Incompatibility	Desirable
Perception	Images	Connectedness, parts	Inconsistency	Seen
Discourse comprehension	Meanings	Semantic relatedness	Inconsistency	Understood
Analogy	Mapping hypotheses	Similarity, structure, purpose	1-1 mappings	Corresponding
Cognitive dissonance	Beliefs, attitudes	Consistency	Inconsistency	Believed
Impression formation	Stereotypes, traits	Association	Negative association	Believed
Democratic deliberation	Actions, goals, propositions	Facilitation, explanation	Incompatible actions and beliefs	Joint action

sense of constraint satisfaction is nontrivial; chapter 8 discusses the alternative claim that inference should be understood probabilistically.

3 COMPUTING COHERENCE

If coherence can indeed be generally characterized in terms of satisfaction of multiple positive and negative problem involving only 100 propositions would require considering 2^{100} = 1,267,650,600,228,229,401,496,703, 205,376 different solutions. No computer, and presumably no mind, can be expected to compute coherence in this way except for trivially small cases.

In computer science, a problem is said to be intractable if there is no deterministic polynomial-time solution to it, i.e., if the amount of time required to solve it increases at a faster-than-polynomial rate as the problem grows in size. For intractable problems, the amount of time and memory space required to solve the problem increases rapidly as the problem size grows. Consider, for example, the problem of using a truth table to check whether a compound proposition is consistent. A proposition with n connectives requires a truth table with 2^n rows. If n is small, there is no difficulty, but an exponentially increasing number of rows is required as n gets larger. Problems in the class NP include ones that can be solved in polynomial time by a *nondeterministic* algorithm that allows guessing.

Members of an important class of problems called NP-complete are equivalent to each other in the sense that if one of them has a polynomial-time solution, then so do all the others. A new problem can be shown to be NP-complete by showing (a) that it can be solved in polynomial time by a nondeterministic algorithm, and (b) that a problem already known to be NP-complete can be transformed into it, so that a polynomial-time solution to the new problem would serve to generate a polynomial-time solution to all the other problems. If only (b) is satisfied, then the problem is said to be NP-hard, i.e., at least as hard as the NP-complete problems. In the past two decades, many problems have been shown to be NP-complete, and deterministic polynomial-time solutions have been found for none of them, so it is widely believed that the NP-

i. Take an arbitrary ordering of the elements $e_1, \ldots, e_n$ of E.
ii. Let A and R, the accepted and rejected elements, be empty.
iii. For each element e_i in the ordering, if adding e_i to A increases the total weight of satisfied constraints more than adding it to R, then add e_i to A; otherwise, add e_i to R.

The problem with this algorithm is that it is seriously dependent on the ordering of the elements. Suppose that we have just 4 elements with a negative constraint between e_1 and e_2 and positive constraints between e_1 and e_3, e_1 and e_4, and e_2 and e_4. In terms of explanatory coherence, e_1 and e_2 could be thought of as competing hypotheses, with e_1 explaining more than e_2, as shown in figure 2.1. The three other algorithms for computing coherence discussed in this section accept e_1, e_3, and e_4, while rejecting e_2. But the serial algorithm will accept e_2 if it happens to come first in the ordering. In general, the serial algorithm does not do as well as the other algorithms at satisfying constraints and accepting the appropriate elements.

Although the serial algorithm is not prescriptively attractive as an account of how coherence should be

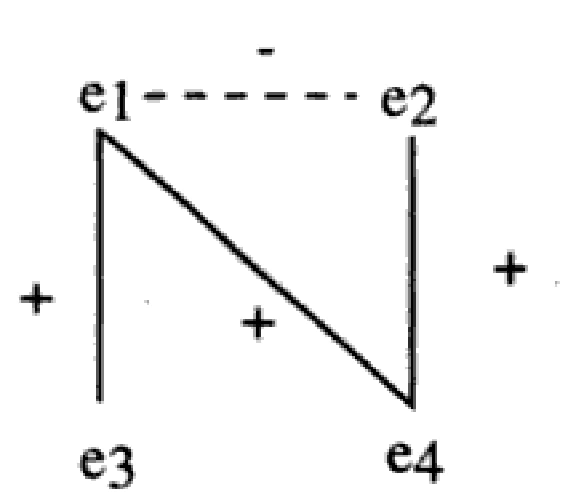

Figure 2.1
A simple coherence problem. Positive constraints are represented by solid lines, and the negative constraint is represented by a dashed line.

computed, it may well describe to some extent people's limited rationality. Ideally, a coherence inference should be nonmonotonic in that maximizing coherence can lead to rejecting elements that were previously accepted. In practice, however, limitations of attention and memory may lead people to adopt local, suboptimal methods for calculating coherence (Hoadley, Ranney, and Schank 1994). Psychological experiments are needed to determine the extent to which people do coherence calculations suboptimally. In general, coherence theories are intended to be both descriptive and prescriptive, in that they describe how people make inferences when they are in accord with the best practices compatible with their cognitive capacities (Thagard 1992b, 97).

Algorithm 3: Connectionist

A more effective method for computing coherence uses connectionist (neural network) algorithms. This method is a generalization of methods that have been successfully applied in computational models of explanatory coherence, deliberative coherence, and elsewhere.

Here is how to translate a coherence problem into a problem that can be solved in a connectionist network:

1. For every element e_i of E, construct a unit u_i that is a node in a network of units U. Such networks are very roughly analogous to networks of neurons.

2. For every positive constraint in $C+$ on elements e_i and e_j, construct a symmetric excitatory link between the corresponding units u_i and u_j. Elements whose acceptance is favored can be positively linked to a special unit whose activation is clamped at the maximum value. Reasons for favoring some classes of elements are discussed in section 7 of chapter 3.

3. For every negative constraint in $C-$ on elements e_i and e_j, construct a symmetric inhibitory link between the corresponding units u_i and u_j.

4. Assign each unit u_i an equal initial activation (say 0.01), then update the activation of all the units in parallel. The updated activation of unit is calculated on the basis of its current activation, the weights on links to other units, and the activation of the units to which it is linked. A number of equations are available for specifying how this updating is done (McClelland and Rumelhart 1989). For example, on each cycle the activation of unit j, a_j, can be updated according to the following equation:

$$a_j(t+1) = a_j(t)(1-d) + \text{net}_j(\max - a_j(t)) \text{ if } \text{net}_j > 0,$$
$$\text{otherwise } \text{net}_j(a_j(t) - \min)$$

Here d is a decay parameter (say 0.05) that decrements each unit at every cycle, min is a minimum activation (−1), max is maximum activation (1). Based on weight w_{ij} between each unit i and j, we can calculate net_j, the net input to a unit, by $\text{net}_j = \Sigma_i w_{ij} a_i(t)$. Although all links in coherence networks are symmetrical, the flow of activation is not, because a special unit with activation clamped at the maximum value spreads activation to favored units linked to it, such as units representing evidence in the explanatory coherence model ECHO. Typically, activation is constrained to remain between a minimum (e.g., −1) and a maximum (e.g., 1).

5. Continue the updating of activation until all units have settled, that is, achieved unchanging activation values. If a unit u_i has final activation above a specified threshold (e.g., 0), then the element e_i represented by u_i is deemed to be accepted. Otherwise, e_i is rejected.

We thus get a partitioning of elements of E into accepted and rejected sets by virtue of the network U set-

Table 2.2
Comparison of coherence problems and connectionist networks

Coherence	Connectionist network
Element	Unit
Positive constraint	Excitatory link
Negative constraint	Inhibitory link
Conditions on coherence	Parallel updating of activation
Element accepted	Unit activated
Element rejected	Unit deactivated

tling in such a way that some units are activated and others deactivated. Intuitively, this solution is a natural one for coherence problems. Just as we want two coherent elements to be accepted or rejected together, so two units connected by an excitatory link will tend to be activated or deactivated together. Just as we want two incoherent elements to have one that is accepted and the other rejected, so two units connected by an inhibitory link will tend to suppress each other's activation, with one activated and the other deactivated. A solution that enforces the two conditions on maximizing coherence is provided by the parallel update algorithm that adjusts the activation of all units at once on the basis of their links and previous activation values. Table 2.2 compares coherence problems and connectionist networks.

Connectionist algorithms can be thought of as maximizing the "goodness of fit" or "harmony" of the network, defined by $\Sigma_i \Sigma_j w_{ij} a_i(t) a_j(t)$, where w_{ij} is the weight on the link between two units, and a_i is the activation of a unit (Rumelhart, Smolensky, Hinton, and McClelland 1986, 13). The characterization of coherence given in section 1 is an abstraction from the notion of goodness of fit. The value of this abstraction is that it provides a general account of well behaved, and they also produce the results that one would expect for coherence maximization. For example, when ARCS is used to retrieve an analog for a representation of *West Side Story* from a data base of representations of 25 of Shakespeare's plays, it retrieves *Romeo and Juliet.*

The dozen coherence problems summarized in table 2.1 might give the impression that the different kinds of inference involved in all the problems occur in isolation from each other. But any general theory of coherence must be able to say how different kinds of coherence can interact. For example, the problem of other minds can be understood as involving both explanatory coherence and analogical coherence: the plausibility of my hypothesis that you have a mind is based both on it being the best explanation of your behavior and on the analogy between your behavior and my behavior (chapter 4, section 4). The interconnections between different kinds of coherence can be effectively modeled by introducing new kinds of constraints between the elements of the different coherence problems. In the problem of other minds, the explanatory-coherence element representing the hypothesis that you have a mind can be connected by a positive constraint with the analogical-coherence element representing the mapping hypothesis that you are similar to me. Choosing the best explanation and the best analogy can then occur simultaneously as interconnected coherence processes. Similarly, ethical justification and epistemic justification can be intertwined through constraints that connect ethical principles and empirical beliefs, for example, about human nature (chap. 5). A full, applied coherence theory would specify the kinds of connecting constraints that interrelate the different kinds of coherence problems. The parallel connectionist algorithm for maximizing coherence has no difficulty in performing the simultaneous evaluation of interconnected coherence problems.

simply by accepting and rejecting hypotheses that represent the best mappings, but by choosing as the best mappings hypotheses represented by units with higher activations than alternative hypotheses. In general, the output of the greedy algorithm, dividing elements into accepted or rejected, is less informative than the output of the connectionist algorithm, which produces activations that indicate *degrees* of acceptance and rejection. Empirical tests of coherence theories have found strong correlations between experimental measurements of people's confidence about explanations and stereotypes and the activation levels produced by connectionist models (Read and Marcus-Newhall 1993, Kunda and Thagard 1996, Schank and Ranney 1992). Hence the connectionist algorithm is much more suitable than the greedy algorithm for modeling psychological data. Moreover, with its use of random solutions and a great many coherence calculations, the greedy algorithm seems less psychologically plausible than the connectionist algorithm.

Algorithm 5: Semidefinite programming

The proof that the graph-theory problem MAX CUT can be transformed to the coherence problem shows a close relation between them (see the appendix to Thagard and Verbeurgt 1998). MAX CUT is a difficult problem in graph theory that until recently had no good approximation: for twenty years the only known approximation technique was one similar to the incremental algorithm for coherence described above. This technique only guarantees an expected value of 0.5 times the optimal value. Recently, however, Goemans and Williamson (1995) discovered an approximation algorithm for MAX CUT that delivers an expected value of at least 0.87856 times the optimal value. Their algorithm depends on rounding

Coherence in Thought and Action
Paul Thagard

This book is an essay on how people make sense of each other and the world they live in. Making sense is the activity of fitting something puzzling into a coherent pattern of mental representations that include concepts, beliefs, goals, and actions. Paul Thagard proposes a general theory of coherence as the satisfaction of multiple interacting constraints, and he discusses the theory's numerous psychological and philosophical applications. Much of human cognition can be understood in terms of coherence as constraint satisfaction, and many of the central problems of philosophy can be given coherence-based solutions. Thagard shows how coherence can help to unify issues in psychology and philosophy, particularly when one is addressing questions of epistemology, metaphysics, ethics, politics, and aesthetics. He also shows how coherence can integrate cognition and emotion.

Paul Thagard is Professor of Philosophy and Director of the Cognitive Science Program at the University of Waterloo. His books include *Mind: Introduction to Cognitive Science* (MIT Press, 1996) and *How Scientists Explain Disease* (1999).

"If there was ever an attempt to construct a grand unified theory of human consciousness with its multifarious manifestations, *Coherence in Thought and Action* is certainly a competing, and capable, candidate."
Metapsychology

Life and Mind: Philosophical Issues in Biology and Psychology series

A Bradford Book

Cover art: the subjective Necker cube. From Bradley and Petry, "Organizational Determinants of Subjective Contour: The Subjective Necker Cube," *American Journal of Psychology* 90 (1977): 253–262.

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

0-262-70092-1

9 780262 700924 90000