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.

On the other hand, if two elements are negatively constrained, we want one to be accepted and the other rejected. Note that these two conditions are intended as desirable results, not as strict requisites of coherence: the partition is intended to maximize compliance with them, not necessarily to ensure that *all* the constraints are simultaneously satisfied, since simultaneous satisfaction may be impossible. The partition is coherent to the extent that A includes elements that cohere with each other while excluding ones that do not cohere with those elements. We can define the *coherence* of a partition of E into A and R as W, the sum of the weights of the constraints on E that satisfy the above two conditions. Coherence is maximized if there is no other partition that has greater total weight.

This abstract characterization applies to the main philosophical and psychological discussions of coherence. It will not handle nonpairwise inconsistencies or incompatibilities, for example, when there is a joint inconsistency among the three propositions "Al is taller than Bob," "Bob is taller than Cary," and "Cary is taller than Al." However, there are computational methods for converting constraint satisfaction problems whose constraints involve more than two elements into binary problems (Bacchus and van Beek 1998). Hence my characterization of coherence in terms of constraints between two elements suffices in principle for dealing with more complex coherence problems with nonbinary constraints.

An unrelated notion of coherence is used in probabilistic accounts of belief, where degrees of belief in a set of propositions are called coherent if they satisfy the axioms of probability (see chapter 8 for a discussion of the relation between coherence and probability). The characterization of coherence as constraint satisfaction does not by itself furnish a way of understanding degrees of 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.

The theory of explanatory coherence shows how constraints can be specified for evaluating hypotheses and other propositions (see Thagard 1989, 1992b, and chap. 3 below). In that theory, positive constraints arise from relations of explanation and analogy that hold between propositions, and negative constraints arise either because two hypotheses contradict each other or because they compete with each other to explain the same evidence.

Russell has argued that the justification of mathematical axioms is similarly a matter of coherence (Russell 1973, see also Kitcher 1983, and chapter 3). Axioms are accepted not because they are a priori true, but because they serve to generate and systematize interesting theorems, which are themselves justified in part because they follow from the axioms.

Goodman contended that the process of justification of logical rules is a matter of making mutual adjustments between rules and accepted inferences, bringing them into conformity with each other (Goodman 1965, Thagard 1988, chap. 7). Logical justification can then be seen as a coherence problem: the elements are logical rules and accepted inferences; the positive constraints derive from justification relations that hold between particular rules and accepted inferences; and the negative constraints arise because some rules and inferences are inconsistent with each other.

Similarly, Rawls (1971) argued that ethical principles can be revised and accepted on the basis of their fit with particular ethical judgments. Determining fit is achieved by adjusting principles and judgments until a balance between them, reflective equilibrium, is achieved. Daniels (1979) advocated that *wide* reflective equilibrium should also require taking into account relevant empirical background theories. Brink (1989) defended a theory of ethical justification based on coherence between moral theories and 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 constraints, we can precisely address the question of how coherence can be computed, i.e., how elements can be selectively accepted or rejected in a way that maximizes compliance with the two coherence conditions on constraint satisfaction. This section describes five algorithms for maximizing coherence:

- An *exhaustive* search algorithm that considers all possible solutions
- An *incremental* algorithm that considers elements in arbitrary order
- A *connectionist* algorithm that uses an artificial neural network to assess coherence
- A *greedy* algorithm that uses locally optimal choices to approximate a globally optimal solution
- A *semidefinite programming* (SDP) algorithm that is guaranteed to satisfy a high proportion of the maximum satisfiable constraints

The first two algorithms are of limited use, but the others provide effective means of computing coherence.

Algorithm 1: Exhaustive

The obvious way to maximize coherence is to consider all the different ways of accepting and rejecting elements. Here is the exhaustive algorithm:

1. Generate all possible ways of partitioning elements into accepted and rejected.
2. Evaluate each of these for the extent to which it achieves coherence.
3. Pick the one with highest value of W.

The problem with this approach is that for n elements, there are 2^n possible acceptance sets. A small coherence

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-complete problems are inherently intractable. (For a review of NP-completeness, see Garey and Johnson 1979; for an account of why computer scientists believe that P ≠ NP, see Thagard 1993.)

Millgram (1991) noticed that the problem of computing coherence appears similar to other problems known to be intractable and conjectured that the coherence problem is also intractable. He was right: Karsten Verbeurgt proved that MAX CUT, a problem in graph theory known to be NP-complete, can be transformed to the coherence problem (Thagard and Verbeurgt 1998, appendix). If there were a polynomial-time solution to coherence maximization, there would also be a polynomial-time solution to MAX CUT and all the other NP-complete problems. So, on the widely held assumption that P ≠ NP (i.e., that the class of problems solvable in polynomial time is not equal to NP), we can conclude that the general problem of computing coherence is computationally intractable. As the number of elements increases, a general solution to the problem of maximizing coherence will presumably require an exponentially increasing amount of time.

For epistemic coherence and any other kind of problem that involves large numbers of elements, this result is potentially disturbing. Each person has thousands or millions of beliefs. Epistemic coherentism requires that justified beliefs must be shown to be coherent with other beliefs. But the transformation of MAX CUT to the coherence problem shows, on the assumption that P ≠ NP, that computing coherence will be an exponentially increasing function of the number of beliefs.

Algorithm 2: Incremental

Here is a simple, efficient serial algorithm for computing coherence:

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

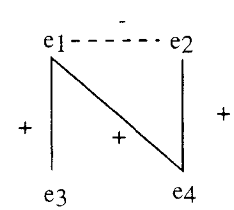


Figure 2.1
A simple coherence problem. Positive constraints are represented by solid lines, and the negative constraint is represented by a dashed line.

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:

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

coherence independent of neural network implementations and makes possible investigation of alternative algorithmic solutions to coherence problems. (See section 4 for discussion of various measures of coherence.)

Despite the natural alignment between coherence problems and connectionist networks, the connectionist algorithms do not provide a universal, guaranteed way of maximizing coherence. We cannot prove in general that connectionist updating maximizes the two conditions on satisfying positive and negative constraints, since settling may achieve only a local maximum. Moreover, there is no guarantee that a given network will settle at all, let alone that it will settle in a number of cycles that is a polynomial function of the number of units.

While there are no mathematical guarantees on the quality of solutions produced by neural networks, empirical results for numerous connectionist models of coherence yield excellent results. ECHO is a computational model of explanatory coherence that has been applied to many cases from the history of science and legal reasoning, including cases with more than 150 propositions (Thagard 1989, 1991, 1992a, 1992b, Nowak and Thagard 1992a, 1992b, Eliasmith and Thagard 1997). Computational experiments have revealed that the number of cycles of activation updating required for settling does not increase as networks become larger: fewer than 200 cycles suffice for all ECHO networks tried so far. ARCS is a computational model of analog retrieval that selects a stored analog from memory on the basis of its having the most coherent match with a given analog (Thagard, Holyoak, Nelson, and Gochfeld 1990). ARCS networks tend to be much larger than ECHO networks—up to more than 400 units and more than 10,000 links—but they still settle in fewer than 200 cycles, and the number of cycles for settling barely increases with network size. Thus, quantitatively these networks are very 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.

Algorithm 4: Greedy

Other algorithms are also available for solving coherence problems efficiently. I owe to Toby Donaldson an algorithm that starts with a randomly generated solution and then improves it by repeatedly flipping elements from the accepted set to the rejected set or vice versa. In computer science, a *greedy* algorithm is one that solves an optimization problem by making a locally optimal choice intended to lead to a globally optimal solution. Selman, Levesque, and Mitchell (1992) presented a greedy algorithm for solving satisfiability problems, and a similar technique produces the following coherence algorithm:

1. Randomly assign the elements of E into A or R.

2. For each element e in E, calculate the gain (or loss) in the weight of satisfied constraints that would result from flipping e, i.e., moving it from A to R if it is in A or moving it from R to A otherwise.

3. Produce a new solution by flipping the element that most increases coherence, i.e., move it from A to R or from R to A. In case of ties, choose randomly.

4. Repeat (2) and (3) until either a maximum number of tries have taken place or until there is no flip that increases coherence.

On the examples on which it has been tested, this algorithm usually produces the same acceptances and rejections as the connectionist algorithm; exceptions arise from the random character of the initial assignment in step 1 and from the greedy algorithm's breaking ties randomly.

Although the greedy algorithm largely replicates the performance of ECHO and DECO on the examples on which we have tried it, it does not replicate the performance of ACME, which does analogical mapping not 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

a solution to a relaxation of a nonlinear optimization problem, which can be formulated as a semidefinite programming (SDP) problem, a generalization of linear programming to semidefinite matrices. Mathematical details and proofs are provided in the appendix to Thagard and Verbeurgt 1998.

From the perspective of coherence, two results are important, one theoretical and the other experimental. Verbeurgt proved that the semidefinite programming technique applied to MAX CUT can also be used for the coherence problem, with the same 0.878 performance guarantee: using this technique guarantees that the weight of the constraints satisfied by a partition into accepted and rejected will be at least 0.878 of the optimal weight. But where does this leave the connectionist algorithm, which has no similar performance guarantee? We have run computational experiments to compare the results of the SDP algorithm to those produced by the connectionist algorithms used in existing programs for explanatory and deliberative coherence. Like the greedy algorithm, the semidefinite-programming solution handles ties between equally coherent partitions differently from the connectionist algorithm, but otherwise it yields equivalent results.

4 MEASURING COHERENCE

The formal constraint-satisfaction characterization of coherence and the various algorithms for computing coherence suggest various means by which coherence can be measured. Such measurement is useful for both philosophical and psychological purposes. Philosophers concerned with normative judgments about the justification of belief systems naturally ask questions about the degree of coherence of a belief or set of beliefs. Psychologists can use the degree of coherence as a variable to correlate with experimental measures of mental performance, such as expressed confidence of judgments.

There are three sorts of measurement of coherence that are potentially useful:

- The degree of coherence of an entire set of elements
- The degree of coherence of a subset of the elements
- The degree of coherence of a particular element

The goodness-of-fit (harmony) measure of a neural network defined in section 3, $\Sigma_i w_{ij} a_i(t) a_j(t)$, can be interpreted as the coherence of an entire set of elements, and the assigned activation values as representing their acceptance and rejection. This measure is of limited use, however, since it is very sensitive to the number of elements, as well as to the particular equations used to update activation in the networks. Sensitivity to the sizes of networks can be overcome by dividing goodness-of-fit by the number of elements or by the number of links or constraints (see Shultz and Lepper 1996). Holyoak and Thagard (1989) found that goodness-of-fit did not give a reliable metric of the degree of difficulty of analogical mapping, which they instead measured in terms of the number of cycles required for a network to settle.

Network-independent measures of coherence can be stated in terms of the definition of a coherence problem given in section 1. For any partition of the set of elements into accepted and rejected, there is a measure *W* of the sum of the weights of the satisfied constraints. Let *W*_opt be the coherence of the optimal solution. The ideal measure of coherence achieved by a particular solution would be *W*/*W*_opt, the ratio of the coherence *W* of the solution to the coherence *W*_opt of the optimal solution; thus the best

solution would have measure one. This measure is difficult to obtain, however, since the value of the optimal solution is not generally known. Another possible measure of coherence is the ratio W/W^*, where W^* is the sum of the weights of all constraints. This ratio does not necessarily indicate the closeness to the optimal solution as *W*/*W*_opt would, but it does have the property that the higher the ratio, the closer the solution is to optimal. Thus it gives a size-independent measure of coherence. In addition, when there is a solution where all constraints are satisfied, W/W^* is equal to *W*/*W*_opt.

Neither goodness-of-fit nor W/W^* provides a way of defining the degree of coherence of a subset of elements. This is unfortunate, since we would like be able to quantify judgments such as "Darwin's theory of evolution is more coherent than creationism," where Darwin's theory consists of a number of hypotheses. The connectionist algorithm does provide a useful way to measure the degree of coherence of a particular element, since the activation of a unit represents the degree of acceptability of the element. The coherence of a set of elements can then be roughly measured as the mean activation of those elements. It would be desirable to define, within the abstract model of coherence as constraint satisfaction, a measure of the degree of coherence of a particular element or of a subset of elements, but it is not clear how to do so. Such coherence is highly nonlinear, since the coherence of an element depends on the coherence of all the elements that constrain it, including elements with which it competes. The coherence of a set of elements is not simply the sum of the weights of the constraints satisfied by accepting them, but depends also on the comparative degree of constraint satisfaction of other elements that negatively constrain them.

5 SUMMARY

Unlike most of the rest of the book, this chapter has been rather technical, in order to provide a rigorous account of coherence. Computing coherence is a matter of maximizing constraint satisfaction and can be accomplished approximately by several different algorithms. The most psychologically appealing models of coherence optimization are provided by connectionist algorithms. These use neuronlike units to represent elements, and excitatory and inhibitory links to represent positive and negative constraints. Settling a connectionist network by spreading activation results in the activation (acceptance) of some units and the deactivation (rejection) of others. Coherence can be measured in terms of the degree of constraint satisfaction accomplished by the various algorithms.

3 *Knowledge*

Many contemporary philosophers favor coherence theories of knowledge (Bender 1989, BonJour 1985, Davidson 1986, Harman 1986, Lehrer 1990). But the nature of coherence is usually left vague, with no method provided for determining whether a belief should be accepted or rejected on the basis of its coherence or incoherence with other beliefs. Haack's (1993) explication of coherence relies largely on an analogy between epistemic justification and crossword puzzles. This chapter shows how epistemic coherence can be understood in terms of maximization of constraint satisfaction, in keeping with the computational theory presented in chapter 2. Knowledge involves at least five different kinds of coherence—explanatory, analogical, deductive, perceptual, and conceptual—each requiring different sorts of elements and constraints.

Explanatory coherence subsumes Susan Haack's recent "foundherentist" theory of knowledge. This chapter shows how her crossword-puzzle analogy for epistemic justification can be interpreted in terms of explanatory coherence and describes how her use of the analogy can be understood in terms of analogical coherence. I then give an account of deductive coherence, showing how the selection of mathematical axioms can be understood as a constraint-satisfaction problem. Moreover, visual interpretation can also be understood in terms of satisfaction of multiple constraints. After a brief account of how conceptual coherence can also be understood in terms of constraint satisfaction, I conclude with a discussion of how the "multicoherence" theory of knowledge avoids many criticisms traditionally made against coherentism.

1 HAACK'S "FOUNDHERENTISM" AND EXPLANATORY COHERENCE

Susan Haack's book *Evidence and Inquiry* (1993) presents a compelling synthesis of foundationalist and coherentist epistemologies. From coherentism, she incorporates the insights that there are no indubitable truths and that beliefs are justified by the extent to which they fit with other beliefs. From empiricist foundationalism, she incorporates the insights that not all beliefs make an equal contribution to the justification of beliefs and that sense experience deserves a special, if not completely privileged, role. She summarizes her "foundherentist" view with the following two principles (Haack 1993, 19):

(FH1) A subject's experience is relevant to the justification of his empirical beliefs, but there need be no privileged class of empirical beliefs justified exclusively by the support of experience, independently of the support of other beliefs.

(FH2) Justification is not exclusively one-directional, but involves pervasive relations of mutual support.

Haack's explication of "pervasive relations of mutual support" relies largely on an analogy with how crossword puzzles are solved by fitting together clues and possible interlocking solutions.

To show that Haack's epistemology can be subsumed within the account of coherence as constraint satisfaction,

I will reinterpret her principles in terms of the theory of explanatory coherence (TEC) and describe how crossword puzzles can be solved as a constraint-satisfaction problem by the computational model (ECHO) that instantiates TEC. TEC is informally stated in the following principles (Thagard 1989, 1992a, 1992b):

Principle E1: Symmetry Explanatory coherence is a symmetric relation, unlike, say, conditional probability. That is, two propositions *p* and *q* cohere with each other equally.

Principle E2: Explanation (a) A hypothesis coheres with what it explains, which can either be evidence or another hypothesis. (b) Hypotheses that together explain some other proposition cohere with each other. (c) The more hypotheses it takes to explain something, the lower the degree of coherence.

Principle E3: Analogy Similar hypotheses that explain similar pieces of evidence cohere.

Principle E4: Data Priority Propositions that describe the results of observations have a degree of acceptability on their own.

Principle E5: Contradiction Contradictory propositions are incoherent with each other.

Principle E6: Competition If *p* and *q* both explain a proposition, and if *p* and *q* are not explanatorily connected, then *p* and *q* are incoherent with each other (*p* and *q* are explanatorily connected if one explains the other or if together they explain something).

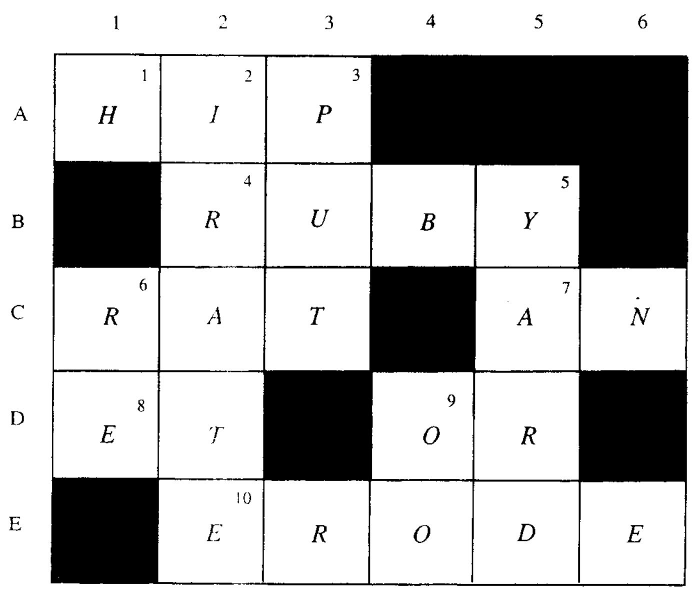
Principle E7: Acceptance The acceptability of a proposition in a system of propositions depends on its coherence with them.

The last principle, Acceptance, states the fundamental assumption of coherence theories that propositions are accepted on the basis of how well they cohere with other propositions. It corresponds to Haack's principle FH2 that acceptance depends not on any deductive derivation but on relations of mutual support. Principle E4, Data Priority, makes it clear that TEC is not a pure coherence theory that treats all propositions equally in the assessment of coherence but, like Haack's principle FH1, gives a certain priority to experience. Like Haack's theory, TEC does not treat sense experience as the source of given, indubitable beliefs, but allows the results of observation and experiment to be overridden on the basis of coherence considerations. For this reason, it is preferable to treat TEC as affirming a kind of discriminating coherentism rather than as a hybrid of coherentism and foundationalism (see the discussion of indiscriminateness in section 7).

TEC goes beyond Haack's foundherentism in specifying more fully the nature of the coherence relations. Principle E2, Explanation, describes how coherence arises from explanatory relations: when hypotheses explain a piece of evidence, the hypotheses cohere with the evidence and with each other. These coherence relations establish the positive constraints required for the global assessment of coherence in line with the characterization of coherence in chapter 2. When a hypothesis explains evidence, this establishes a positive constraint that tends to make them either accepted together or rejected together. In some cases, evidence can also contribute to the explanation, as when a hypothesis in conjunction with observations explains some other observation. Then the hypothesis and the evidence used in the explanation cohere on the basis of statement (b) of principle E2 rather than statement (a).

Principle E3, Analogy, establishes positive constraints between hypotheses that accomplish similar explanations. The negative constraints required for a global assessment of coherence are established by principles E5 and E6, Contradiction and Competition. When two propositions are incoherent with each other because they are contradictory or in explanatory competition, there is a negative constraint between them that will tend to make one of them accepted and the other rejected. Principle E4, Data Priority, can also be interpreted in terms of constraints, by positing a special element EVIDENCE that is always accepted and that has positive constraints with all evidence derived from sense experience. The requirement to satisfy as many constraints as possible will tend to lead to the acceptance of all elements that have positive constraints with EVIDENCE, but their acceptance is not guaranteed. Constraints are *soft*, in that coherence maximizing will tend to satisfy them, but not all constraints will be satisfied simultaneously.

As chapter 2 showed, the idea of maximizing constraint satisfaction is sufficiently precise that it can be computed using a variety of algorithms. The theory of explanatory coherence (TEC) is instantiated by a computer program (ECHO) that uses input about explanatory relations and contradiction to create a constraint network that performs the acceptance and rejection of propositions on the basis of their coherence relations. ECHO can be used to simulate the solution of Haack's crossword-puzzle analogy for foundherentism. Figure 3.1 is the example that Haack uses to illustrate how foundherentism envisages mutual support. In the crossword puzzle, the clues are analogous to sense experience and provide a basis for filling in the letters. But the clues are vague and do not themselves establish the entries, which must fit with the other entries. Filling in each entry depends not only on the clue for it but also on how the entries fit with each other. In terms of coherence as constraint satisfaction, we can say that there are positive constraints connecting particular letters with each other and with the clues. For example, in 1 across the hypothesis that the first letter is H coheres with the hypotheses that the second letter is I and the third letter is P. Together, these hypotheses provide an explanation of the clue, since "hip" is the start of the cheerful expression "Hip hip hooray!" Moreover, the hypothesis that I is the



ACROSS

1 A cheerful start (3)
4 She's a jewel (4)
6 No, it's Polonius (3)
8 A visitor from outside fills this space (2)
9 What's the alternative? (2)
10 Dick Turpin did this to York; it wore 'im out (5)

DOWN

2 Angry Irish rebels (5)
3 Have a shot at an Olympic event (3)
5 A measure of one's back garden (4)
6 What's this all about? (2)
9 The printer hasn't got my number (2)

Figure 3.1
Crossword puzzle used to illustrate coherence relations, adapted from Haack 1993 (p. 85).

second letter of 1 across must cohere with the hypothesis that I is the first letter of 2 down, which, along with other hypotheses about the word for 2 down, provides an answer for the clue for 2 down. These coherence relations are positive constraints that a computation of the maximally

coherent interpretation of the crossword puzzle should satisfy. Contradictions can establish incoherence relations: only one letter can fill each square, so if the first letter of 1 across is H, it cannot be another letter.

Chris Eliasmith simulated a solution to the crossword puzzle, using the program ECHO, that takes input of the following form:

(explain (hypothesis 1 hypothesis 2 ...) evidence)

For the crossword puzzle, we can identify each square using a system of letters A to E down the left side and numbers 1 to 6 along the top, so that location of the first letter of 1 across is A1. Then we can write A1 = H to represent the hypothesis that the letter H fills this square. Writing C1a for the clue for 1 across, ECHO can be given the following input:

(explain (A1=H A2=I A3=P) C1a)

This input establishes positive constraints among all pairs of the four elements listed, so that the hypotheses that the letters are H, I, and P tend to be accepted or rejected together in company with the clue C1a. Since the clue is given, it is treated as data, and therefore the element C1a has a positive constraint with the special EVIDENCE element, which is accepted. For real crossword puzzles, explanation is not quite the appropriate relation to describe the connection between entries and clues, but it is appropriate here because Haack uses the crossword-puzzle example to illuminate explanatory reasoning. (A full statement of the input to ECHO to handle the crossword-puzzle example can be found in the appendix to Thagard, Eliasmith, Rusnock, and Shelley, forthcoming, available on the Web at http://cogsci.uwaterloo.ca/articles/pages/epistemic.html.) ECHO does not model how people solve the crossword puzzle by working out clues one at a time, but it does serve to evaluate a full solution as one that is generally coherent.

The crossword-puzzle analogy is useful in showing how beliefs can be accepted or rejected on the basis of how well they fit together. But TEC and ECHO go well beyond the analogy, since they demonstrate how coherence can be computed. ECHO not only has been used to simulate the crossword-puzzle example; it has been applied to many of the most important cases of theory choice in the history of science, as well as to examples from legal reasoning and everyday life (Eliasmith and Thagard 1997; Nowak and Thagard 1992a, 1992b; Thagard 1989, 1992b, 1999). Moreover, ECHO has provided simulations of the results of a variety of experiments in social and educational psychology, so it meshes with a naturalistic approach to epistemology tied with human cognitive processes (Read and Marcus-Newhall 1993, Schank and Ranney 1992, Byrne 1995). Thus the construal of coherence as constraint satisfaction, as manifested in the theory of explanatory coherence and the computational model ECHO, subsumes Haack's foundherentism.

2 ANALOGICAL COHERENCE

Although explanatory coherence is the most important contributor to epistemic justification, it is not the only kind of coherence. While the crossword-puzzle analogy plays a central role in her presentation of foundherentism, Haack nowhere acknowledges the important contributions of analogies to epistemic justification. TEC's principle E3 allows such a contribution, since it establishes coherence (and hence positive constraints) among analogous hypotheses. This principle was based on the frequent use of analogies by scientists, for example, Darwin's use of the

Table 3.1
Analogical mapping between epistemic justification and crossword puzzle completion

Epistemic justification	Crossword puzzles
Observations	Clues
Explanatory hypotheses	Words
Explanatory coherence	Words fitting with clues and each other

analogy between artificial and natural selection in support of his theory of evolution.

Using analogies, as Haack does when she compares epistemic justification to crossword puzzles, requires the ability to map between two analogs, the target problem to be solved and the source that is intended to provide a solution. Mapping between source and target is a difficult computational task, but in recent years a number of computational models have been developed that perform it effectively. Haack's analogy between epistemic justification and crossword puzzles uses the mapping shown in table 3.1.

Analogical mapping can be understood in terms of coherence and multiple constraint satisfaction, where the elements are hypotheses concerning what maps to what and the main constraints are similarity, structure, and purpose (Holyoak and Thagard 1995). To highlight the similarities and differences with explanatory coherence, here are comparable principles of analogical coherence:

Principle A1: Symmetry Analogical coherence is a symmetric relation among mapping hypotheses.

Principle A2: Structure A mapping hypothesis that connects two propositions, $R(a, b)$ and $S(c, d)$, coheres with mapping hypotheses that connect R with S, a with c, and b with d. And all those mapping hypotheses cohere with each other.

Principle A3: Similarity Mapping hypotheses that connect elements that are semantically or visually similar have a degree of acceptability on their own.

Principle A4: Purpose Mapping hypotheses that provide possible contributions to the purpose of the analogy have a degree of acceptability on their own.

Principle A5: Competition Mapping hypotheses that offer different mappings for the same object or concept are incoherent with each other.

Principle A6: Acceptance The acceptability of a mapping hypothesis in a system of mapping hypotheses depends on its coherence with them.

In analogical mapping, the coherence elements are hypotheses concerning which objects and concepts correspond to each other. Initially, mapping favors hypotheses that relate similar objects and concepts (A3). Depending on whether analogs are represented verbally or visually, the relevant kind of similarity is either semantic or visual. For example, when Darwin drew an analogy between natural and artificial selection, both analogs had verbal representations of selection, which had similar meaning. In visual analogies, perceptual similarity can suggest possible correspondences, for example, when the atom with its electrons circling the nucleus is pictorially compared to the solar system with its planets revolving around the sun. We then get the positive constraint that if two objects or concepts in an analogy are visually or semantically similar to each other, then an analogical mapping that puts them in correspondence with each other should tend to be accepted. This kind of similarity is much more local and direct than the more general overall similarity that is found between two analogs. Another positive constraint is pragmatic: we want to encourage mappings that can accomplish the

purposes of the analogy such as problem solving or explanation (A4).

Additional positive constraints arise because of the need for structural consistency (A2). In the verbal representations (CIRCLE (ELECTRON NUCLEUS)) and (REVOLVE (PLANET SUN)), maintaining structure (i.e., keeping the mapping as isomorphic as possible) requires that if we map CIRCLE to REVOLVE, then we must map ELECTRON to PLANET and NUCLEUS to SUN. The need to maintain structure establishes positive constraints, so that, for example, the hypothesis that CIRCLE corresponds to REVOLVE will tend to be accepted with or rejected with the hypothesis that ELECTRON corresponds to PLANET. Negative constraints occur between hypotheses representing incompatible mappings, for example, between, the hypothesis that the atom corresponds to the sun and the hypothesis that the atom corresponds to a planet (A5). Principles A2 and A5 together incline, but do not require, analogical mappings to be isomorphisms. Analogical coherence is a matter of accepting the mapping hypotheses that satisfy the most constraints.

The multiconstraint theory of analogy just sketched has been applied computationally to a great many examples and has provided explanations for numerous psychological phenomena. Also epistemologically important is the fact that the constraint-satisfaction construal of coherence provides a way of unifying explanatory and analogical epistemic issues. Chapter 4 argues that the solution to the philosophical problem of other minds (that is, whether there are any) requires a combination of explanatory and analogical coherence. Thus metaphysics, like science, can employ a combination of explanatory and analogical coherence to defend important conclusions. Mathematical knowledge, however, is more dependent on deductive coherence.

3 DEDUCTIVE COHERENCE

For millennia, epistemology has been enthralled by mathematics, taking mathematical knowledge as the purest and soundest type. The Euclidean model of starting with indubitable axioms and deriving equally indubitable theorems has influenced many generations of philosophers. Surprisingly, however, Bertrand Russell, one of the giants of the axiomatic method in the foundations of mathematics, had a different view of the structure of mathematical knowledge. In an essay he presented in 1907, Russell remarked on the apparent absurdity of proceeding from recondite propositions in symbolic logic to the proof of such truisms as 2 + 2 = 4. He concluded,

> The usual mathematical method of laying down certain premises and proceeding to deduce their consequences, though it is the right method of exposition, does not, except in the more advanced portions, give the order of knowledge. This has been concealed by the fact that the propositions traditionally taken as premises are for the most part very obvious, with the fortunate exception of the axiom of parallels. But when we push the analyses farther, and get to more ultimate premises, the obviousness becomes less, and the analogy with the procedure of other sciences becomes more visible. (Russell 1973, 282)

Just as scientists discover hypotheses from which facts of the senses can be deduced, so mathematicians discover premises (axioms) from which elementary propositions (theorems) such as 2 + 2 = 4 can be derived. Unlike the logical axioms that Russell, following Frege, used to derive arithmetic, these theorems are often intuitively obvious. Russell contrasts the a priori obviousness of such mathematical propositions with the lesser obviousness of the senses, but notes that obviousness is a matter of degree and that even where there is the highest degree of obviousness,

we cannot assume that the propositions are infallible, since they may be abandoned because of conflict with other propositions. Thus for Russell, adoption of a system of mathematical axioms and theorems is much like the scientific process of acceptance of explanatory hypotheses. Let us try to exploit this analogy to develop a theory of deductive coherence.

The elements are mathematical propositions—potential axioms and theorems. The positive and negative constraints can be established by coherence and incoherence relations specified by a set of principles that are adapted from the seven principles of explanatory coherence in section 1.

Principle D1: Symmetry Deductive coherence is a symmetric relation among propositions, unlike, say, deductive entailment.

Principle D2: Deduction (a) An axiom or other proposition coheres with propositions that are deducible from it. (b) Propositions that together are used to deduce some other proposition cohere with each other. (c) The more hypotheses it takes to deduce something, the less the degree of coherence.

Principle D3: Intuitive Priority Propositions that are intuitively obvious have a degree of acceptability on their own. Propositions that are obviously false have a degree of rejectability on their own.

Principle D4: Contradiction Contradictory propositions are incoherent with each other.

Principle D5: Acceptance The acceptability of a proposition in a system of propositions depends on its coherence with them.

When a theorem is deduced from an axiom, the axiom and theorem cohere symmetrically with each other, which allows the theorem to confer support on the axiom as well as vice versa, just as an explanatory hypothesis and the evidence it explains confer support on each other (principles D1, D2). Principle D2, Deduction, is just like the second principle of explanatory coherence, but with the replacement of the coherence-producing relation of explanation by the similarly coherence-producing relation of deduction. These coherence relations are the source of positive constraints: when an axiom and theorem cohere because of the deductive relation between them, there is a positive constraint between them, so that they will tend to be accepted together or rejected together. Statement (c) of the principle has the consequence that the weight of the constraint will be reduced if the deduction requires other propositions. Just as scientists prefer simpler theories, other things being equal, Russell looked for simplicity in axiom systems: "Assuming, then, that elementary arithmetic is true, we may ask for the fewest and simplest logical principles from which it can be deduced" (Russell 1973, 275–276).

Although some explanations are deductive, not all are, and not all deductions are explanatory (Kitcher and Salmon 1989). So explanatory coherence and deductive coherence cannot be assimilated to each other. The explanatory-coherence principle E4, Data Priority, discriminated in favor of the results of sensory observations and experiments, but deductive coherence in mathematics requires a different kind of intuitive obviousness. Russell remarks that the obviousness of propositions such as $2 + 2 = 4$ derives remotely from the empirical obviousness of such observations as that 2 sheep + 2 sheep = 4 sheep. Principle D3, Intuitive Priority, does not address the source of the intuitiveness of mathematical propositions, but simply takes into account that it exists. Different axioms and theorems will have different degrees of intuitive priority. D3 provides discriminating constraints that encourage the acceptance of intuitively obvious propositions such as $2 + 2 = 4$. Russell stressed the need to avoid having falsehoods as consequences of axioms, so I have included in D3 a

specific mention of intuitively obvious falsehoods being rejected, even though it is redundant: a falsehood can be indirectly rejected because it contradicts an obvious truth. Principle D4, Contradiction, establishes negative constraints that prevent two contradictory propositions from being accepted simultaneously. For mathematics, these should be constraints with very high weights. Even in mathematics, however, there is sometimes the need to live with contradictions until a way around them can be found, as when Russell discovered the paradoxes of set theory. The contradiction principle is obvious, but it is much less obvious whether there is competition between mathematical axioms in the same way there is between explanatory hypotheses, so I have not included a competition principle.

Whereas there are ample scientific examples of the role of analogy in enhancing explanatory coherence, cases of an analogical contribution to deductive coherence in mathematics are rarer, so my principles of deductive coherence do not include an analogy principle, although analogy is important in mathematical discovery (Polya 1957). Moreover, analogical considerations can indirectly enter into the choice of mathematical principles by virtue of isomorphisms between areas of mathematics that allow all the theorems in one area to be translated into theorems in the other, as when geometry is translated into Cartesian algebra.

Russell does not explicitly defend a coherentist justification of axiom systems, but he does remark, "We tend to believe the premises because we can see that their consequences are true, instead of believing the consequences because we know the premises to be true" (Russell 1973, 273–274). According to Russell, there are additional noncoherence considerations such as independence and convenience that contribute to selection of an axiom set.

Philip Kitcher (1983, 220) sees the contribution of important axiomatizations by Euclid, Cayley, Zermelo, and Kolmogorov as analogous to the uncontroversial cases in which scientific theories are adopted because of their power to unify. Principle D5, Acceptance, summarizes how axioms can be accepted on the basis of the theorems they yield, while at the same time theorems are accepted on the basis of their derivation from axioms. The propositions to be accepted are just the ones that are most coherent with each other, as shown by finding a partition of propositions into accepted and rejected sets in a way that satisfies the most constraints.

This section has discussed deductive coherence in the context of mathematics, but it is also relevant to other domains such as ethics. According to Rawl's notion of reflective equilibrium, ethical principles such as "Killing is wrong" are to be accepted or rejected on the basis of how well they fit with particular ethical judgments such as "Killing Salman Rushdie is wrong" (Rawls 1971). Ethical coherence is not only deductive coherence, however, since *wide* reflective equilibrium requires finding the most coherent set of principles and particular judgments in the light of background information, which can introduce considerations of explanatory, analogical, and deliberative coherence (chapter 5). Principle D3, Intuitive Priority, is much more problematic for ethics than for mathematics, since there is much greater diversity in ethical intuitions than in mathematical intuitions. Nobody denies that $2 + 2 = 4$, but debates rage concerning such topics as the morality of abortion. (See chapter 5 for further discussion of the role of intuition in coherence-based inference.)

Just as explanatory coherence looks for a good fit between hypotheses and evidence, deductive coherence looks for a good fit between general principles and intuitive judgments. Perception can also be construed as a coherence problem.

4 PERCEPTUAL COHERENCE

Explanatory and deductive coherence both involve propositional elements, but not all knowledge is verbal. Our perceptual knowledge includes visual, auditory, olfactory, and tactile representations of what we see, hear, smell, and feel. According to most current theories, visual perception is not a matter of directly apprehending the world, but requires inference and constraint satisfaction (Rock 1983, Kosslyn 1994). Vision is not simply a matter of taking sensory inputs and transforming them directly into interpretations that form part of conscious experience, because the sensory inputs are often incomplete or ambiguous. For example, the subjective Necker cube in figure 3.2 can be seen in two

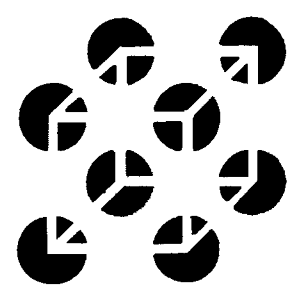


Figure 3.2
The subjective Necker cube. The perceived top edge can be seen either as being at the front or at the back of the cube. Try to make it flip back and forth by concentrating on different edges. (Source: Bradley and Petry 1977, p. 254. Copyright 1977 by Board of Trustees of the University of Illinois. Used with permission of the University of Illinois Press.)

different ways with different front faces. I shall not attempt here anything like a full theory of different kinds of perception, but I want to sketch how vision can be understood as a coherence problem similar to but different from the kinds of coherence so far discussed.

Visual perception begins with two-dimensional image arrays on the retina, but the visual interpretations that constitute sensory experience are much more complex than these arrays. How does the brain construct a coherent understanding of sensory inputs? In visual coherence, the elements are nonverbal representations of input images and full-blown visual interpretations, which fit together in accord with the following principles:

Principle V1: Symmetry Visual coherence is a symmetric relation between a visual interpretation and a low-level representation of sensory input.

Principle V2: Interpretation A visual interpretation coheres with a representation of sensory input if they are connected by perceptual principles such as proximity, similarity, and continuity.

Principle V3: Sensory priority Sensory input representations are acceptable on their own.

Principle V4: Incompatibility Incompatible visual interpretations are incoherent with each other.

Principle V5: Acceptance The acceptability of a visual interpretation depends on its coherence with sensory inputs, other visual interpretations, and background knowledge.

Principle V2, Interpretation, asserts that how an interpretation fits with sensory input is governed by innate perceptual principles such as ones described in the 1930s by Gestalt psychologists (Koffka 1935). According to the principle of proximity, visual parts that are near to each other join together to form patterns or groupings. Thus an interpretation that joins two visual parts together in a pattern will cohere with sensory input that has the two

parts close to each other. According to the Gestalt principle of similarity, visual parts that resemble each other in respect to form size, color, or direction unite to form a homogeneous group. Hence an interpretation that combines resembling parts in a pattern will cohere with sensory input that has parts similar to each other. Other Gestalt principles encourage interpretations that find continuities and closure (lack of gaps) in sensory inputs. The visual system also has built into it assumptions that enable it to use such cues as size constancy, texture gradients, motion parallax, and retinal disparity to provide connections between visual interpretations and sensory inputs (Medin and Ross 1992, chap. 5). These assumptions establish coherence relations between visual interpretations and sensory inputs, and thereby provide positive constraints that tend to make visual interpretations accepted along with the sensory inputs with which they cohere.

Image arrays on the retina are caused by physical processes not subject to cognitive control, so we can take them as given (V3). But considerable processing begins even at the retinal level, and many layers of visual processing occur before a person has a perceptual experience. The sensory inputs may be given, but sensory experience certainly is not. Sensory inputs may fit with multiple possible visual interpretations that are incompatible with each other and are therefore incoherent and the source of negative constraints (V4).

Thus the Gestalt principles and other assumptions built into the human visual system establish coherence relations that provide positive constraints linking visual interpretations with sensory input. Negative constraints arise between incompatible visual interpretations, such as the two ways of seeing the Necker cube. Our overall visual experience arises from accepting the visual interpretation that satisfies the most positive and negative constraints.

Coherence thus produces our visual knowledge, just as it establishes our explanatory and deductive knowledge.

I cannot attempt here to sketch coherence theories of other kinds of perception: smell, sound, taste, touch. Each would have a different version of principle V2, Interpretation, involving its own kinds of coherence relations based on the innate perceptual system for that modality.

5 CONCEPTUAL COHERENCE

Given the above discussions of explanatory, deductive, analogical, and perceptual coherence, the reader might now be worried about the proliferation of kinds of coherence: just how many are there? I see the need to discuss only one additional kind of coherence, conceptual, that seems important for understanding human knowledge.

Different kinds of coherence are distinguished from each other by the different kinds of elements and constraints they involve. In explanatory coherence, the elements are propositions and the constraints are explanation-related, but in *conceptual* coherence the elements are concepts and the constraints are derived from positive and negative associations among concepts. Much work has been done in social psychology to examine how people apply stereotypes when forming impressions of other people. For example, you might be told that someone is a woman pilot who likes monster-truck rallies. Your concepts of *woman, pilot,* and *monster-truck fan* may involve a variety of concordant and discordant associations that need to be reconciled as part of the overall impression you form of this person.

Conceptual coherence can be characterized with principles similar to those already presented for other kinds of coherence:

Principle C1: Symmetry Conceptual coherence is a symmetric relation between pairs of concepts.

Principle C2: Association A concept coheres with another concept if they are positively associated, i.e., if there are objects to which they both apply.

Principle C3: Given Concepts The applicability of a concept to an object, for example, of the concept *woman* to a particular person, may be given perceptually or by some other reliable source.

Principle C4: Negative Association A concept incoheres with another concept if they are negatively associated, i.e., if an object falling under one concept tends not to fall under the other concept.

Principle C5: Acceptance The applicability of a concept to an object depends on the applicability of other concepts.

Taken together, these principles explain how people decide what complexes of concepts apply to a particular object.

The association of concepts can be understood in terms of social stereotypes. For example, the stereotypes that some Americans have of Canadians include associations with other concepts such as *polite, law-abiding, beer-drinking,* and *hockey-playing*, where these concepts have different kinds of associations each other. The stereotype that Canadians are polite (a Canadian is someone who says "Thank you" to bank machines) conflicts with the stereotype that hockey players are somewhat crude. If you are told that someone is a Canadian hockey player, what impression do you form of him? Applying stereotypes in complex situations is a matter of conceptual coherence, where the elements are concepts and the positive and negative constraints are positive and negative associations between concepts (C2, C4). Some concepts cohere with each other (e.g., law-abiding and polite), while other concepts resist cohering with each other (e.g., polite and crude). The applicability of some concepts is given, as when you can see that someone is a hockey player or are told by a reliable source that he or she is a Canadian (C3).

Many psychological phenomena concerning how people apply stereotypes can be explained in terms of conceptual-constraint satisfaction. Kunda and Thagard (1996) were able to account for most of the phenomena emerging from the literature on how people form impressions of others based on stereotypes and individuating information. Their connectionist program, IMP, successfully simulated the results of experiments that demonstrated these phenomena. For example, Kunda, Sinclair, and Griffin (1997) found that the impact of stereotypes on impressions can depend on the perceiver's judgment task and that the effects of stereotypes on trait ratings of an individual were undermined by the individual's behavior. Although construction workers are stereotyped as more aggressive than accountants, a construction worker and an accountant were viewed as equally unaggressive after having failed to react to an insult, an unaggressive behavior. But even though the stereotypes no longer affected trait ratings, they continued to influence predictions about the individual's behavior: the construction worker was still viewed as more likely than the accountant to engage in coarse aggressive behaviors such as punching and cursing.

The parallel constraint-satisfaction model predicts such a pattern when the stereotypes are associated with additional traits that are not undermined by the target's behavior and so can continue to influence behavioral predictions. In this case, even though both targets came to be viewed as equally unaggressive, the construction worker

continued to be viewed as a member of the working class, and the accountant as a member of the upper middle class. Punching and cursing are positively associated with working-class status but negatively associated with upper middle-class status. Therefore, the working-class construction worker was viewed as more likely than the upper middle-class accountant to punch and curse even though the two were viewed as equally unaggressive. Conceptual coherence leads to different inferences.

There are thus five primary kinds of coherence relevant to assessing knowledge: explanatory, analogical, deductive, perceptual, and conceptual. (A sixth kind, deliberative coherence, is relevant to decision making and ethics; it is discussed in chapter 5.) Each kind of coherence involves a set of elements and positive negative constraints, including constraints that discriminate in order to favor the acceptance or rejection of some of the elements, as summarized in table 3.2. A major problem for the kind of multifaceted coherence theory of knowledge I have been presenting concerns how these different kinds of coherence relate to each other. To solve this problem, I would need

Table 3.2
Kinds of coherence and their constraints

	Elements	Positive constraints	Discriminating constraints	Negative constraints
Explanatory	Hypotheses, evidence	E2, explanation E3, analogy	E4, data priority	E5, contradiction E6, competition
Analogical	Mapping hypotheses	A2, structure	A3, similarity A4, purpose	A5, competition
Deductive	Axioms, theorems	D2, deductive entailment	D3, intuitive priority	D4, contradiction
Visual	Visual interpretations	V2, interpretation	V3, sensory priority	V4, incompatibility
Conceptual	Concepts	C2, association	C3, given concepts	C4, negative association

Names such as "E2" refer to principles stated in the text.

to describe in detail the interactions involved in each of the fifteen different pairs of kinds of coherence. Some of these pairs are straightforward. For example, explanatory and deductive coherence both involve propositional elements and very similar kinds of constraints. In addition, chapter 4 shows how explanatory and analogical coherence can interact in the problem of other minds.

The relation, however, between propositional elements (explanatory and deductive) on the one hand and visual and conceptual elements is obscure; it is not obvious, for example, how a system of explanatory coherence can interface with a system of visual coherence. One possibility is that a deeper representational level, such as the systems of vectors used in neural networks, may provided a common substratum for propositional, perceptual, and conceptual coherence.

Note that simplicity plays a role in most kinds of coherence. It is explicit in explanatory and deductive coherence, where an increase in the number of propositions required for an explanation or deduction decreases simplicity, and deliberative coherence is similar. Simplicity is implicit in analogical coherence, which encourages 1-1 mappings. Perhaps simplicity plays a role in perceptual coherence as well (Rock 1983, 146).

6 UNIFYING COHERENCE

This presentation of five kinds of coherence raises some serious questions about whether the list is exclusive and exhaustive. Are these kinds of coherence really different from each other, or are some merely variants? Are there other kinds of coherence important for cognition? How do the different kinds of coherence work together?

Some of the five kinds of coherence are indeed quite similar to each other. Explanatory and deductive coherence are alike in that both involve relations among propositions according to similar principles: compare E1–E7 with D1–D5. But I prefer to keep them distinct as separate kinds of coherence because of important differences between their fundamental coherence relations and the associated principles. Deductive coherence is based on purely deductive relations between propositions, as for example when "All cities have roads" implies "Toronto has roads." In contrast, although explanation may sometimes involve deduction, as in theories in mathematical physics, it is fundamentally a matter of there being a causal relation between what is explained and the representations that do the explaining (see Thagard 1999, chap. 7, for a defense of this view of explanation). Moreover, the source of priority is different in the two kinds of coherence. In explanatory coherence, priority is given to propositions that describe the results of experience and observation (principle E4), whereas in deductive coherence, priority accrues to propositions such as $2 + 2 = 4$, whose obviousness may rest on reasoning as well as observation (principle D3).

Conceptual coherence might seem a lot like explanatory or deductive coherence, since for inferential purposes concepts can be translated into propositions. Instead of activating the concept *pilot* to indicate that it applies to a woman Mary, we could speak instead of activating the proposition "Mary is a pilot." To do so, however, would be to obscure the direct connections of positive and negative association that exist between concepts, for example, between *pilot* and *male* and *daring*. There is abundant experimental and computational evidence that concepts are a psychologically realistic kind of mental representation not reducible to propositions (Thagard 1996, chap. 4). Moreover, the associative relations between concepts are much looser than the explanatory and deductive relations required for those kinds of coherence, so the constraints between elements in conceptual coherence deserve to be treated separately.

How many other kinds of coherence are there? Chapter 5 discusses deliberative coherence, which concerns how decisions are made on the basis of coherence among actions and goals. This sixth kind of coherence is, as far as I know, the only additional one needed to cover the main kinds of inference that people perform. Deliberative coherence concerns inferences about what to do, so it is not discussed in this chapter, which concerns inferences relevant to the development of knowledge. Chapter 6 discusses emotional coherence, which is not, however, a seventh kind of coherence along the lines so far discussed. Rather, it provides an expanded way of considering the elements and constraints of the six basic kinds of coherence, by adding emotional attitudes toward the elements.

Having six kinds of coherence might suggest that inference is a confused jumble, but they in fact suggest a unified view of coherence-based inference. All six kinds of coherence are specified in terms of elements and constraints, and we saw in chapter 2 that there are algorithms for maximizing constraint satisfaction. Hence once constraints and elements are specified, the same inference engine can work to decide which elements to accept and which to reject. The only rule of inference is this: accept a conclusion if its acceptance maximizes coherence. Different kinds of coherence furnish different kinds of elements connected by different kinds of constraints, but inference is performed by exactly the same kind of constraint-satisfaction algorithm working simultaneously with all the different elements and

constraints. This makes possible a unified account of inferences based on more than one kind of coherence. Later chapters provide extended examples of complex inferences involving mixtures of explanatory, analogical, and other kinds of coherence (see chap. 4 on the problem of other minds, chap. 5 on capital punishment, and chap. 6 on trust).

Although much of cognition can be understood in terms of coherence mechanisms, there is obviously more to cognition than achieving coherence among a set of given elements. Cognition is also generative, producing new concepts, propositions, and analogies. Moreover, for coherence to be assessed, constraints among elements need to have been generated.

Generation of new elements is sometimes driven by *in*coherence. If I am trying to understand someone but fail to form a coherent impression or attribution, I may be spurred to form new elements that can add coherence to the old set of elements. To take an example from Kunda et al. 1990, if I am told that someone is a Harvard-educated carpenter, it may be difficult to reconcile the conflicting expectations associated with the two concepts. Surprise is an emotional reaction that signals that a satisfactory degree of coherence has not been achieved (see chapter 6). This reaction triggers hypothesis formation, as I ask myself how someone with a Harvard degree could end up working as a carpenter. People show ingenuity in generating explanations, for example, that the Harvard graduate was a counterculture type who preferred a nonprofessional career path. Hence new hypotheses and possibly also new concepts (*Ivy League laborer*) can be added to the set of elements so as to lend greater coherence to the attempt to make sense of this person. In this case, generation of elements is incoherence-driven: it is prompted by a failure to achieve an interpretation that satisfies an adequate number of the positive and negative constraints. In addition to surprise, other emotions, such as anxiety, may also signal incoherence.

Not all element generation is incoherence-driven, however. Some representations arise serendipitously, based on things we just happen to encounter. I may form a concept of Albanians as the result of meeting various immigrants from Albania, without having experienced any incoherence in my previous attempts to understand them. In other cases, new representations may arise from curiosity-driven thinking that is motivated not by any incoherence but by the desire to find out more about something that interests me. If I am interested in the Balkans, I will learn more about Serbs and Croats and may form stereotypes about them without having tried and failed to fit them together with my other social concepts. Motivation may also lead one to generate new concepts. For example, our desire to protect our stereotypes from change in the face of disconfirmation may lead us to assign individuals who threaten our stereotypes into novel subtypes that serve to isolate these individuals from their group (Kunda and Oleson 1995). Thus serendipity, curiosity, and motivation, in addition to incoherence, can spur the generation of new representations.

Where do constraints come from? Some may be innate, capturing basic conceptual relations such as that an object cannot be both red and black all over. Most constraints, however, capture empirically discovered relations between elements. For conceptual coherence, I learn that some concepts (e.g., nurse and benevolent) are positively associated, whereas other concepts (e.g., Nazi and benevolent) are negatively associated. Such associations may be learned through direct observation of nurses or Nazis as well as through cultural transmission. For explanatory coherence, the positive constraints come from understand-

ing causal relations. The link between the hypothesis that Mary is in love and the fact to be explained that Mary is very happy depends on the causal judgment gleaned from experience that being in love can cause people to be happy. Negative constraints in explanatory coherence arise from logical contradictions (you cannot be both in love and not in love) and from competing hypotheses (maybe instead she's happy because she got a promotion at work).

Because any full account of human cognition would have to include an account of how new concepts, hypotheses, and other representations are formed, a complete cognitive architecture would have to include generation mechanisms as well as coherence mechanisms (see Thagard 1996 for a review of different kinds of learning). My goal in this book is not to propose a cognitive architecture, but merely to show how coherence mechanisms contribute to making sense of people and events.

Explanatory, analogical, deductive, visual, and conceptual coherence add up to a comprehensive, computable, naturalistic theory of epistemic coherence. Let us now see how this theory can handle some of the standard objections that have been made to coherentist epistemologies.

7 OBJECTIONS TO COHERENCE THEORIES

Vagueness

One common objection to coherence theories is *vagueness*: in contrast to fully specified theories of deductive and inductive inference, coherence theories have generally been vague about what coherence is and how coherent elements can be selected. My general characterization of coherence shows how vagueness can be overcome. First, for a particular kind of coherence, it is necessary to specify the nature of the elements and define the positive and negative constraints that hold between them. This task has been accomplished for the kinds of coherence discussed above. Second, once the elements and constraints have been specified, it is possible to use connectionist algorithms to compute coherence, accepting and rejecting elements in a way that approximately maximizes compliance with the coherence conditions (chapter 2). Computing coherence can then be as exact as deduction or probabilistic reasoning (chapter 8), and can avoid the problems of computational intractability that arise with them. Being able to do this computation does not, of course, help with the problem of generating elements and constraints, but it does show how to make a judgment of coherence with the elements and constraints on hand. Arriving at a rich, coherent set of elements—scientific theories, ethical principles, or whatever—is a very complex process that intermingles both assessment of coherence and generation of new elements; the parallel constraint-satisfaction algorithm shows only how to do the first of these. Whether a cognitive task can be construed as a coherence problem depends on the extent to which it involves evaluation of how existing elements fit together rather than generation of new elements.

Indiscriminateness

The second objection to coherence theories is *indiscriminateness*: coherence theories fail to allow that some kinds of information deserve to be treated more seriously than others. For example, in epistemic justification, it has been argued that perceptual beliefs should be taken more seriously in determining general coherence than mere speculation. The abstract characterization of coherence given in

chapter 2 is indiscriminating, in that all elements are treated equally in determinations of coherence.

But all the kinds of coherence discussed above are discriminating in the sense of allowing favored elements of E to be given priority in being chosen for the set of accepted elements A. We can define a discriminating-coherence problem as one where members of a subset D of E are favored to be members of A. Favoring them does not guarantee that they will be accepted: if there were such a guarantee, the problem would be foundationalist rather than coherentist, and D would constitute the foundation for all other elements. As Audi (1993) points out, even foundationalists face a coherence problem in trying to decide what beliefs to accept in addition to the foundational ones. Explanatory coherence treats hypothesis evaluation as a discriminating-coherence problem, since it gives priority to propositions that describe observational and experimental results. That theory is not foundationalist, since evidential propositions can be rejected if they fail to cohere with the entire set of propositions. Similarly, table 3.2 makes it clear that the other five kinds of coherence are also discriminating.

Computing a solution to a discriminating-coherence problem involves only a small addition to the characterization of coherence given in chapter 2, p. 18:

> For each element d in the discriminated set D, construct a positive constraint between d and a special element e_s that is assigned to the set A of accepted elements.

The effect of having a special element that constrains members of the set D is that the favored elements will tend to be accepted, without any guarantee that they will be accepted. Chapter 2 already described how the connectionist algorithm for coherence implements the discrimination condition by having an excitatory link between the unit representing d and a special unit that has a fixed, unchanging maximum activation (i.e., 1). The effect of constructing such links to a special unit is that when activation is updated, it flows directly from the activated special unit to the units representing the discriminated elements. Hence those units will more strongly tend to end up activated than nondiscriminated ones and will have a greater effect on which other units get activated. The algorithm does not, however, enforce the activation of units representing discriminated elements, which can be deactivated if they have strong inhibitory links with other activated elements. Thus a coherence computation can be discriminating while remaining coherentist.

We can thus distinguish between three kinds of coherence problems. A *pure* coherence problem is one that does not favor any elements as potentially worthy of acceptance. A *foundational* coherence problem selects a set of favored elements for acceptance as self-justified. A *discriminating* coherence problem favors a set of elements but their acceptance still depends on their coherence with all the other elements. I have shown how coherence algorithms can naturally treat problems as discriminating without being foundational.

Isolation

The *isolation* objection has been characterized as follows:

> This objection states that the coherence of a theory is an inadequate justification of the theory, because by itself it doesn't supply the necessary criteria to distinguish it from illusory but consistent theories. Fairytales may sometimes be coherent as may dreams and hallucinations. Astrology may be as coherent as astronomy, Newtonian physics as coherent as Einsteinian physics. (Pojman 1993, 191)

Thus an isolated set of beliefs may be internally coherent but should not be judged to be justified.

My characterization of coherence provides two ways of overcoming the isolation objection. First, as we just saw, a coherence problem may be discriminating, giving nonabsolute priority to empirical evidence or other elements that are known to make a relatively reliable contribution to solution of the kind of problem at hand. The comparative coherence of astronomy and astrology is thus in part a matter of coherence with empirical evidence, of which there is obviously far more for astronomy than astrology. Second, the existence of negative constraints such as inconsistency shows that we cannot treat astronomy and astrology as isolated bodies of beliefs. The explanations of human behavior offered by astrology often conflict with those offered by psychological science. Astrology might be taken to be coherent on its own, but once it offers explanations that compete with psychology and astronomy, it becomes a strong candidate for rejection. The isolation objection may be a problem for underspecified coherence theories that lack discrimination and negative constraints, but it is easily overcome by the constraint-satisfaction approach.

Having negative constraints, however, does not guarantee consistency in the accepted set A. The second coherence condition, which encourages dividing negatively constrained elements between A and R, is not rigid, so there may be cases where two negatively constrained elements both end up being accepted. For a correspondence theory of truth, this is a disaster, since two contradictory propositions cannot both be true. It would probably also be unappealing to most advocates of a coherence theory of truth. To overcome the consistency problem, we could revise the second coherence condition by making it rigid: a partition of elements (propositions) into accepted and rejected sets must be such that if e_i and e_j are inconsistent, then if e_i is in A then e_j *must* be in R. I do not want, however, to defend a coherence theory of truth, since there are good reasons for preferring a correspondence theory based on scientific realism (chapter 4).

For a coherence theory of epistemic justification, inconsistency in the set A of accepted propositions is also problematic, but we can leave open the possibility that coherence is temporarily maximized by adopting an inconsistent set of beliefs. We might deal with the lottery and proofreading paradoxes simply by being inconsistent, believing that a lottery is fair while believing of each ticket that it will not win, or believing that a paper must have a typographical error in it somewhere while believing of each sentence that it is flawless. A more interesting case is the relation between quantum theory and general relativity, two theories that individually possess enormous explanatory coherence. According to the eminent mathematical physicist Edward Witten, "The basic problem in modern physics is that these two pillars are incompatible. If you try to combine gravity with quantum mechanics, you find that you get nonsense from a mathematical point of view. You write down formulae which ought to be quantum gravitational formulae and you get all kinds of infinities" (Davies and Brown 1988, 90). Quantum theory and general relativity may be incompatible, but it would be folly, given their independent evidential support, to suppose that one must be rejected. Another inconsistency in current astrophysics derives from measurements that suggest that the stars are older than the universe. But astrophysics carries on, just as mathematics did when Russell discovered that Frege's axioms for arithmetic lead to contradictions.

From the perspective of formal logic, contradictions are disastrous, since from any proposition and its negation any formula can be derived: from p to p *or* q by addition, then from not p to q by disjunctive syllogism. Logicians who have wanted to deal with inconsistencies have been forced to resort to relevance or paraconsistent logics. But from the perspective of a coherence theory of inference, there is no need for any special logic. It may turn out at a particular time that coherence is maximized by accepting a set A that is inconsistent, but other coherence-based inferences need not be unduly influenced by the inconsistency, whose effects may be relatively isolated in the network of elements.

Conservatism

Coherence theories of justification may seem unduly conservative in that they require new elements to fit into an existing coherent structure. This charge is legitimate against serial coherence algorithms that determine for each new element whether accepting it increases coherence or not. The connectionist algorithm in chapter 2, on the other hand, allows a new element to enter into the full-blown computation of coherence maximization. If units have already settled into a stable activation, it will be difficult for a new element with no activation to dislodge the accepted ones, so the network will exhibit a modest conservatism. But if new elements are sufficiently coherent with other elements, they can dislodge previously accepted ones. Connectionist networks can be used to model the dramatic shifts in explanatory coherence that take place in scientific revolutions (Thagard 1992b).

Circularity

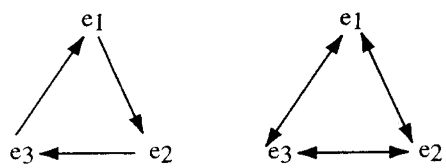
Another standard objection to coherence theories is that they are circular, licensing the inference of p from q and then of q from p. Logic books warn against the fallacy of begging the question, in which someone argues in a circle to infer something from itself. A typical example is someone who argues that God exists because it says so in the Bible, and that you can trust the Bible because its writing was inspired by God. Such circular arguments obviously fail to prove anything, and at first glance coherence-based inference seems circular, since many propositions may serve to support each other.

The theories of coherence and the coherence algorithms presented here make it clear that coherence-based inferences are very different from those familiar from deductive logic, where propositions are derived from other propositions in linear fashion. The characterization of coherence and the algorithms for computing it (chapter 2) involve a global, parallel, but effective means of assessing a whole set of elements simultaneously on the basis of their mutual dependencies. Inference can be seen to be holistic in a way that is nonmystical, computationally effective, and psychologically and neurologically plausible (pairs of real neurons do not excite each other symmetrically, but neuronal groups can). Deductive circular reasoning is inherently defective, but the foundational view that conceives of knowledge as building deductively on indubitable axioms is not even supportable in mathematics, as we saw in section 3. Inference based on coherence judgments is not circular in the way feared by logicians, since it effectively calculates how a whole set of elements fit together, without linear inference of p from q and then of q from p.

Coherentists such as Bosanquet (1920) and BonJour (1985) denied that the circularity evident in coherence-based justification is vicious, and the algorithms for computing coherence in chapter 2 show more precisely how a set of elements can depend on each other interactively, rather than serially. Using the connectionist algorithm, we

can say that after a network of units has settled and some units are identified as being activated, then acceptance of each element represented by an activated unit is justified on the basis of its relation to all other elements. The algorithms for determining activation (acceptance) proceed fully in parallel, with each unit's activation depending on the activation of all connected units after the previous cycle. Because it is clear how the activation of each unit depends simultaneously on the activation of all other units, there need be no mystery about how acceptance can be the result of mutual dependencies. Similarly, the greedy and SDP algorithms in chapter 2 maximize constraint satisfaction globally, not by evaluating individual elements sequentially. Thus modern models of computation vindicate Bosanquet's claim that inference need not be interpreted within the confines of linear systems of logical inference.

Coherence-based inference involves no regress because it proceeds not in steps but rather by simultaneous evaluation of multiple elements. Figure 3.3a shows a viciously circular pattern of inference that starts with e_1, then infers e_2, then infers e_3, then argues in a circle back to e_1. In contrast, figure 3.3b shows the situation when a connectionist algorithm computes everything at once. Unlike entailment or conditional probability, coherence constraints are symmetric relations, which is represented by the double-headed arrows in figure 3.3b. The two-headed arrows indicate that the elements are mutually interdependent, not that one is to be inferred from the others. Activation flows mutually between the elements, but in a realistic example of inference there will also be elements that have inhibitory links with e_1 or some other elements, and some elements will be favored and given a degree of priority. The result is a pattern of inference that looks nothing at all like the circular reasoning in figure 3.3a.



a *b*

Figure 3.3
Circular versus noncircular justification. On the left (a) is a circular series of linear inferences. On the right (b) is a set of elements that mutually support each other.

Truth

Coherence-based reasoning is thus not circular, but it is still legitimate to ask whether it is effective. Do inferences based on explanatory and other kinds of coherence produce true conclusions? Early proponents of coherence theories of inference such as Blanshard (1939) also advocated a coherence theory of truth, according to which the truth of a proposition is constituted by its being part of a general coherent set of propositions. From the perspective of a coherence theory of truth, it is trivial to say that coherence-based inference produces true (i.e., coherent) conclusions. But a major problem arises when we try to justify coherence-based inference with respect to a correspondence theory of truth, according to which the truth of a proposition is constituted by its relation to an external, mind-independent world.

Proponents of coherence theories of truth reject the idea of such an independent world, but considerations of explanatory coherence strongly support its existence, as I argue in chapter 4. Hence truth is a matter also of correspondence, not coherence alone. The issue of correspondence is most acute for pure coherence problems, in which

acceptance of elements is based only on their relation to each other. But the coherence theories that have so far been computationally implemented all treat coherence problems as discriminating. For example, explanatory coherence theory gives priority (but not guaranteed acceptance) to elements representing the results of observation and experiment. Connectionist algorithms naturally implement this discrimination by spreading activation first to units representing elements that should be favored in the coherence calculation; then the activation of other units depends heavily on the activation of those initially activated units. For example, in the explanatory coherence program ECHO, activation spreads first to units representing observational elements, giving them a degree of priority even though they may eventually be rejected on the basis of the overall coherence calculation. Therefore, if we assume with the correspondence theory of truth that observation and experiment involve in part causal interaction with the world, we can have some confidence that the hypotheses adopted on the basis of explanatory coherence also correspond to the world and are not mere mental contrivances that are only internally coherent.

Given a correspondence theory of truth and the consistency of the world, a contradictory set of propositions cannot all be true. But no one ever suggested that coherentist methods guarantee the avoidance of falsehood. All that we can expect of epistemic coherence is that it is generally reliable in accepting the true and rejecting the false. Scientific thinking based on explanatory and analogical coherence has produced theories with substantial technological application, intersubjective agreement, and cumulativity. Our visual systems are subject to occasional illusions, but these are rare compared with the great preponderance of visual interpretations that enable us successfully to interact with the world. Not surprisingly, there is no foundational justification of coherentism, only the coherentist justification that coherentist principles fit well with what we believe and what we do. Temporary tolerance of contradictions may be a useful strategy in accomplishing the long-term aim of accepting many true propositions and few false ones. Hence there is no incompatibility between my account of epistemic coherence and a correspondence theory of truth.

The problem of correspondence to the world is even more serious for ethical justification, for it is not obvious how to legitimate a coherence-based ethical judgment such as "It is permissible to eat some animals but not people." Chapter 5 argues that ethical coherence involves complex interactions of deliberative, deductive, analogical, and explanatory coherence. In some cases the relative objectivity of explanatory coherence, discriminating as it does in favor of observation and experiment, can carry over to the objectivity of ethical judgments that also involve other kinds of coherence. We will see, however, that achieving rational consensus in ethics is more problematic than in epistemology.

8 LANGUAGE

My general account of coherence as constraint satisfaction and the five kinds of coherence discussed in this chapter have many potential applications for understanding people's knowledge and use of language. The topic of linguistic coherence deserves a chapter or even a volume of its own, but I have neither the expertise nor the inclination to discuss it at length. Instead, this section will merely provide pointers to some of the vast literature on coherence in language, along with brief suggestions

concerning how linguistic phenomena can be viewed from the perspective of coherence as constraint satisfaction.

Part of the process of making sense of spoken and written language is dealing with syntactic ambiguities, as in the sentence "I saw her duck," in which "duck" can be either a noun or a verb. Spivey-Knowlton, Trueswell, and Tanenhaus (1993) argue for a constraint-based approach to parsing, in which syntactically relevant contextual constraints provide evidence for or against competing alternatives. Similarly, Menzel (1998) views parsing as a procedure of structural disambiguation that can be modeled by constraint-satisfaction techniques. Prince and Smolensky (1997) discuss phonological grammar in terms of optimizing the satisfaction of multiple constraints on representational well-formedness. These research programs suggest at least the possibility of construing syntactic and phonological interpretation as coherence problems of the sort defined in chapter 2.

Semantic ambiguity can also be handled by constraint-satisfaction methods. Cottrell (1988) proposes a connectionist model of lexical access in which alternative interpretations of an ambiguous word such as "deck" can be evaluated by representing alternative meanings (e.g., ship floor, pack of cards) as nodes in a constraint network. An algorithm similar to the connectionist one described in chapter 2 suffices to activate or deactivate the nodes in accord with how well they fit a given context. Similarly, as chapter 2 described, Kintsch (1988) models comprehension of ambiguous words such as "bank" in terms of associative nets, with one interpretation of the word connected to "river" and another interpretation connected to "money." Spreading excitatory and inhibitory activation enables the nets to select the most appropriate meaning for the context. Semantic disambiguation along these lines is a case of conceptual coherence as described earlier in this chapter.

In his superb book *Comprehension*, Kintsch (1998) applies his construction-integration model to many other processes involved in understanding text, including inference, memory, and problem solving. He describes the integration phase as "essentially a constraint satisfaction process that rejects inappropriate local constructions in favor of those that fit together into a coherent whole" (1998, 119). It is straightforward to interpret his account of integration in terms of the theory of coherence presented in chapter 2. Other discussions of text comprehension that can be brought within the purview of the theory of coherence as constraint satisfaction include Trabasso and Suh's (1993) account of the relevance of explanatory coherence and van den Broek's (1994) analysis of the role of causal and anaphoric relations. Analogical coherence is also relevant to comprehension of texts involving metaphor (Holyoak and Thagard 1995).

However, pursuing linguistic applications of the theory of coherence as constraint satisfaction would take me too far afield from the philosophical and psychological issues concerning inference that are my main concern. This section does not pretend to provide a theory of linguistic coherence, but should help direct anyone interested in constructing one to some of the relevant ingredients.

9 SUMMARY

This chapter has described knowledge in terms of five contributory kinds of coherence: explanatory, analogical, deductive, visual, and conceptual. By analogy to previously presented principles of explanatory coherence, it has generated new principles to capture existing theories of analogical and conceptual coherence, and it has developed new theories of deductive and visual coherence. All of these

kinds of coherence can be construed in terms of constraint satisfaction and computed using connectionist and other algorithms. Haack's "foundherentist" epistemology can be subsumed within the more precise framework offered here, and many of the standard philosophical objections to coherentism can be answered within this framework. The theory of coherence also has applications to the psychological processes by which people understand discourse and other people.