# COHERENCE: THE PRICE IS RIGHT

### Paul Thagard

ABSTRACT: This article is a response to Elijah Millgram's argument that my characterization of coherence as constraint satisfaction is inadequate for philosophical purposes because it provides no guarantee that the most coherent theory available will be true. I argue that the constraint satisfaction account of coherence satisfies the philosophical, computational, and psychological prerequisites for the development of epistemological and ethical theories.

As Elijah Millgram (2000) reports, the idea of coherence has received widespread use in many areas of philosophy. He argues that coherentist approaches suffer from a woeful lack of specification of what coherence is and of how we can tell whether one theory in science, or ethics, or everyday life is more coherent than another. The exception he recognizes is the computational treatment of coherence problems that Karsten Verbeurgt and I have developed (Thagard and Verbeurgt 1998). On this account, a coherence problem consists of a set of elements connected by positive and negative constraints, and a solution consists of partitioning the elements into two sets (accepted and rejected) in a way that maximizes satisfaction of the constraints. Algorithms have been developed that efficiently compute coherence by maximizing constraint satisfaction.

Millgram, however, doubts that this characterization of coherence is fully adequate for philosophical purposes. His main objection is that it is not

**Paul Thagard** is Professor of Philosophy (with cross appointment to Psychology and Computer Science), Director of the Cognitive Science Program, and University Research Chair at the University of Waterloo, Canada. He is the author of *The Cognitive Science of Science: Explanation, Discovery, and Conceptual Change* (MIT Press, 2012); *The Brain and the Meaning of Life* (Princeton University Press, 2010); *Hot Thought: Mechanisms and Applications of Emotional Cognition* (MIT Press, 2006); *Coherence in Thought and Action* (MIT Press, 2000); *How Scientists Explain Disease* (Princeton University Press, 1999); *Mind: Introduction to Cognitive Science* (MIT Press, 1996); and *Conceptual Revolutions* (Princeton University Press, 1992). He is co-author of *Mental Leaps: Analogy in Creative Thought* (with Keith Holyoak, MIT Press, 1995) and *Induction: Processes of Inference, Learning, and Discovery* (with John Holland, Keith Holyoak, and Richard Nisbett, MIT Press, 1986). He is also editor of *Philosophy of Psychology and Cognitive Science* (Elsevier, 2007).

appropriate for epistemology because it provides no guarantee that the most coherent theory available will be true. He contends that if a philosopher wants to invoke coherence, the price of the ticket is to provide a specification that is as concrete and precise as the constraint satisfaction one, but he surmises that alternative characterizations will have flaws similar to the ones he thinks he has identified.

I will argue that the constraint satisfaction account of coherence is not at all flawed in the ways that Millgram describes and, in fact, satisfies the philosophical, computational, and psychological prerequisites for the development of epistemological and ethical theories. Not only have Verbeurgt and I paid the price of the ticket, it is the right price.

First, it is necessary to clear up some confusion about the relation between the general characterization of coherence and the more particular coherence theories that can be applied to philosophical problems. Verbeurgt and I gave the following specification of the general problem of coherence.

> COHERENCE: Let E be a finite set of elements $\{e_i\}$ and let 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. Each constraint is associated with a number $w$, which is the weight (strength) of the constraint. The problem is to partition E into two sets, A (accepted) and R (rejected), in a way that maximizes compliance with the following two *coherence conditions*:
>
> (1) If $(e_i, e_j)$ is in C+, then $e_i$ is in A if and only if $e_j$ is in A.
>
> (2) 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.

By itself, this characterization has no philosophical or psychological applications because it does not state the nature of the elements, the nature of the constraints, or the algorithms to be used to maximize satisfaction of the constraints. In *Coherence in Thought and Action* (2000), I propose that there are six main kinds of coherence: explanatory, deductive, conceptual, analogical, perceptual, and deliberative, each with its own array of elements and constraints. Once these elements and constraints are specified, the algorithms that solve the general coherence problem can be used to compute coherence in ways that apply to philosophical problems. Epistemic coherence is a combination of the first five kinds of coherence, and ethics involves deliberative coherence as well.

Millgram's main objection to coherence as constraint satisfaction concerns the acceptance of scientific theories, which is largely a matter of explanatory coherence (Thagard 1992). The theory of explanatory coherence is stated informally by the following principles:

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

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

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

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

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

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.)

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

I will not elucidate these principles here but merely note how they add content to the abstract specification of coherence as constraint satisfaction. For explanatory coherence, the elements are propositions, and the negative constraints are between propositions that are incoherent with each other, either because they contradict each other or because they compete to explain some other proposition. Positive constraints involve propositions that cohere with each other, linking hypotheses with evidence and also tying together hypotheses that jointly explain some other proposition. In addition, there is a positive constraint encouraging but not demanding the acceptance of propositions describing the results of observation. Explanatory coherence can be computed using a variety of algorithms that maximize constraint satisfaction, including ones using artificial neural networks.

Now we can consider Millgram's argument that "coherentism does not make a whole lot of sense as an approach to science" (2000, 90). To evaluate this claim, we need not merely to look at the abstract characterization of coherence as constraint satisfaction but also at the theory of explanatory coherence that shows how competing explanatory theories can be evaluated. Millgram's worry is this: Verbeurgt has proven that the general problem of coherence is computationally intractable, so the algorithms for computing coherence are approximations; hence, the algorithms do not guarantee that we will get the most coherent theory, and the theory that is *really* the most coherent might be very different from the one chosen by the algorithms.

That this is not a fatal objection to explanatory coherence is evident from the fact that it would apply to any serious attempt to understand scientific inference. Consider the view that scientific theory choice is based not on explanatory coherence but on Bayesian reasoning.[1] On this view, the best theory is the one with the highest subjective probability, given the evidence as calculated by Bayes's theorem. It is well known that the general problem with Bayesian inference is that it is computationally intractable, so the algorithms used for computing posterior probabilities have to be approximations (Cooper 1990; Pearl 1988). Hence, the Bayesian algorithms do not guarantee that we will get the most probable theory, and the theory that is *really* the most probable might be very different from the one chosen by the algorithms. If Millgram's argument is effective against the explanatory coherence account of theory choice, it would also be effective against the Bayesian account and against any other method based on approximation.

Millgram could naturally reply that this argument just shows that Bayesianism is as epistemologically problematic as coherentism and so should also be rejected. But he is implicitly throwing out any procedure rich enough to model scientific theory choice. A model of inference that evaluates competing explanations in a much simpler manner than the Bayesian and coherentist models has been shown to be computationally intractable (Bylander et al. 1991), so we can be confident that *any* algorithm that evaluates competing scientific theories will involve some approximation to picking the best theory. Naturally, this adds to the uncertainty about whether the theory is true, but uncertainty is inherent in any kind of nondeductive inference.

What kind of method would come with a guarantee that it produces true theories? Here are two options.

*Rationalist*: Start with only *a priori* true propositions and deductively derive their consequences.

---

[1] See, e.g., Howson and Urbach 1989 and Maher 1993.

*Empiricist*: Start with only indubitable sense data and believe only what
follows directly from them.

Both of these methods would guarantee truth, but we know that neither can
even begin to account for scientific knowledge. Millgram's argument is not a
criticism of a particular theory of coherence, but it is rather an unsatisfiable
demand to avoid the inherent uncertainty that accompanies nondeductive
inference.

Truth is not the only aim of science; science also values explanatory
unification and practical application. Consider the following propositions.

(1)  Species evolve by natural selection.

(2)  Lemurs have toes.

Both of these propositions are true, but the first is far more important to
science because, as Darwin himself stated, it explains a great number of
previously unconnected facts. It is obvious that explanatory coherence fur-
thers the adoption of theories that encourage explanatory unification because
its principle of data priority, E4, encourages the acceptance of observation
propositions, and its principle of explanation, E2, encourages the acceptance
of hypotheses that explain many observations.

Does explanatory coherence also further the adoption of *true* theories? I
think the answer is "yes," but the argument has to be indirect.

(P1)  Scientists have achieved theories that are at least approximately true.

(P2)  Scientists use explanatory coherence.

(C )  So explanatory coherence leads to theories that are at least approxi-
      mately true.

This is not the place to defend (P1) and (P2), which I have done at length
elsewhere (Thagard 1988, 1992, 1999). (P1) is supported primarily by the
technological applicability of theories in the natural sciences, and (P2) is
supported by computational modeling of many important historical cases of
scientific reasoning. We should not expect a direct defense of the truth
achievement of any nontrivial inductive method, so the indirectness of the
defense of explanatory coherence as a generator of scientific truths does not
undermine coherentism. More recently, I (2007) have argued that we can
judge that a scientific theory is progressively approximating the truth if it
increases its explanatory coherence in two key respects: broadening by explain-
ing more phenomena and deepening by investigating layers of mechanisms.

Let me now address several of Millgram's minor objections to coherence
construed as constraint satisfaction. He contends (2000, 85) that COHER-

ENCE is insensitive to the internal structure of scientific theories, but he ignores the relevant principle of explanatory coherence. According to principle E2(b), hypotheses that together explain a proposition cohere with each other, so there is a positive constraint between them. Hence, computing coherence using algorithms for maximizing constraint satisfaction will encourage that they be accepted together or rejected together, enabling the coherence calculation to be sensitive to its internal structure. Millgram's mistake is misconstruing COHERENCE as a general theory of inference, rather than as the mathematical scaffolding that needs to be filled in with an account of the nature of the elements and constraints that are relevant to inference of a particular sort.

Suppose we have two competing theories, T1 consisting of hypotheses H1 and H2, and T2 consisting of hypotheses H3 and H4. Both theories explain two pieces of evidence, E1 and E2, but T2 provides a more unified explanation: H1 explains E1 on its own, and H2 explains E2 on its own; but H3 and H4 together explain E1, and together explain E2. Intuitively, T2 is more coherent than T1 because of its tighter internal structure. The formal characterization of coherence as constraint satisfaction does not in itself provide guidance about which theory to accept because accepting the first nonunified theory satisfies as many constraints as accepting the second unified theory. The program ECHO, which uses artificial neural network algorithms to maximize coherence, does prefer T2 to T1 because spreading activation through the network of nodes produces a kind of resonance between H3 and H4 that enables them to become more activated than their competitors H1 and H2. Whether the other algorithms that maximize coherence more directly behave in similar fashion depends on the impact of the simplicity principle, E2(c), according to which the more hypotheses it takes to explain something, the lower the degree of coherence. If the degree of coherence between each of two hypotheses that explain a piece of evidence is exactly half the degree of coherence between a hypothesis and a piece of evidence that it explains all by itself, then the coherence algorithms show no preference for internal structure. But if the degree of coherence is reduced by a factor less than the number of hypotheses that together do the explaining, then the coherence algorithms take this into account and prefer theories such as T2 with more internal structure.

Millgram states that "approximating an index of coherence—the weight of the best partition—does not mean approximating the best partition" (2000, 89). Mathematically he is correct, but practically we can acquire evidence that coherence algorithms, including ones proven to approximate an index, do in fact approximate the best partition. This is what computational experiments are for. We can run different coherence algorithms on many different

examples, large and small, and see whether they yield partitions that seem to
be reasonable, given the input. Explanatory coherence has been evaluated for
many important cases in the history of science by seeing whether it captures
the judgments of important scientists. Given the explanatory relations recog-
nized by scientists such as Copernicus, Newton, Lavoisier, and Darwin, a
simulation of explanatory coherence should yield the same judgment as the
scientist; and it does. The simulations show that the theory of explanatory
coherence is coherent with important cases in the history of science.

For complex theories, the demand to simulate the reasoning of scientists is
computationally nontrivial. Chris Eliasmith and I did an explanatory-
coherence analysis of the acceptance of the wave theory of light and chal-
lenged advocates of Bayesian inference to produce an analysis that is as
historically detailed and computationally feasible (Eliasmith and Thagard
1997). As far as I know, this challenge has not been met, and serious practical
limitations on Bayesian inference make me doubt that it can be (Thagard
2000, ch. 8; 2004).

There are good reasons to believe that a coherentist approach that
employs both the abstract characterization of coherence as constraint satis-
faction and its concrete instantiation in terms of principles for explanatory
coherence is adequate for understanding scientific theory choice. Millgram's
arguments do not undermine coherentist epistemology: the price of the ticket
for invoking coherence has already been paid.[2]

## REFERENCES

Bylander, T., D. Allemang, M. Tanner, and J. Josephson. 1991. The computational
    complexity of abduction. *Artificial Intelligence* 49: 25–60.
Cooper, G. 1990. The computational complexity of probabilistic inference using Bayesian
    belief networks. *Artificial Intelligence* 42: 393–405.
Eliasmith, C., and P. Thagard. 1997. Waves, particles, and explanatory coherence. *British
    Journal for the Philosophy of Science* 48: 1–19.
Howson, C., and P. Urbach. 1989. *Scientific reasoning: The Bayesian tradition*. Lasalle, IL:
    Open Court.
Maher, P. 1993. *Betting on theories*. Cambridge: Cambridge University Press.
Millgram, E. 2000. Coherence: The price of the ticket. *Journal of Philosophy* 97: 82–93.
Pearl, J. 1988. *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufman.
Thagard, P. 1988. *Computational philosophy of science*. Cambridge, MA: MIT Press.
———. 1992. *Conceptual revolutions*. Princeton: Princeton University Press.
———. 1999. *How scientists explain disease*. Princeton: Princeton University Press.

———. 2000. *Coherence in thought and action.* Cambridge, MA: MIT Press.

———. 2004. Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence* 18: 231–49.

———. 2007. Coherence, truth, and the development of scientific knowledge. *Philosophy of Science* 74: 28–47.

Thagard, P., and K. Verbeurgt. 1998. Coherence as constraint satisfaction. *Cognitive Science* 22: 1–24.