

COHERENCE, TRUTH, AND THE DEVELOPMENT OF SCIENTIFIC KNOWLEDGE

Paul Thagard
University of Waterloo
pthagard@uwaterloo.ca

Thagard, P. (forthcoming). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*.

Abstract. What is the relation between coherence and truth? This paper rejects numerous answers to this question, including: truth is coherence; coherence is irrelevant to truth; coherence always leads to truth; coherence leads to probability which leads to truth. I will argue that coherence of the right kind leads to at least approximate truth. The right kind is explanatory coherence, where explanation consists of describing mechanisms. We can judge that a scientific theory is progressively approximating the truth if it is increasing its explanatory coherence in two key respects: broadening by explaining more phenomena, and deepening by investigating layers of mechanisms. I sketch an explanation of why deepening is a good epistemic strategy, and discuss the prospect of deepening knowledge in the social sciences and everyday life.

1. Introduction

The problem of the relation between coherence and truth is important for philosophy of science and for epistemology in general. Many epistemologists maintain that epistemic claims are justified, not by a priori or empirical foundations, but by assessing whether they are part of the most coherent account (see, for example, Bonjour, 1985; Harman, 1986; Lehrer, 1990). A major issue for coherentist epistemology concerns whether we are ever warranted in concluding that the most coherent account is true. In the philosophy of science, the problem of coherence and truth is part of the

August 31, 2006

ongoing controversy about scientific realism, the view that science aims at and to some extent succeeds in achieving true theories. The history of science is replete with highly coherent theories that have turned out to be false, which may suggest that coherence with empirical evidence is a poor guide to truth.

This paper argues for a more optimistic conclusion, that coherence of the right kind leads to approximate truth. The right kind is explanatory coherence that involves theories that progressively broaden and deepen over time, where broadening is explanation of new facts and deepening is explanation of why the theory works. First, however, I will consider alternative accounts of the relation between coherence and truth, including the claims that coherence is truth, that coherence is irrelevant to truth, and that probability theory provides the link between coherence and truth.

I take coherence to be a relation among mental representations, including sentence-like propositions and also word-like concepts and picture-like images. Coherence is a global relation among a whole set of representations, but arises from relations of coherence and incoherence between pairs of representations. Section 3 describes how this works in detail for explanatory coherence. As a preliminary account of truth, let me offer the following, adapted from Goldman (1999, p. 59): a representation such as a proposition is true if and only if it purports to describe reality and its content fits reality.

The theory I will develop about the relation of coherence and truth is naturalistic in two respects. First, my theory of coherence is psychologistic in that it employs a model of how human minds make inferences based on coherence considerations. Second, my main conclusion about how coherence can lead to truth is based on examples

from the history of science, under the assumption that natural science is the major source of human knowledge. This paper is not *about* naturalistic epistemology, but is an instance of it.

2. The Relation Between Coherence and Truth

Before developing my own proposal for when coherence leads to truth, I shall deal very briefly with several other accounts about the relation of coherence to truth. The most audacious of these is the coherence theory of truth, according to which the truth of a representation consists in its coherence with other representations, not in its correspondence to a non-mental world; advocates include Blanshard (1939) and Rescher (1973). There are many standard objections to the coherence theory of truth (Young, 2001), but here I mention only a novel one. Coherence with scientific evidence strongly suggests that the universe is more than 10 billion years old, but that representations constructed by humans have existed for less than a million. Thus we can infer that there was a world existing independent of any human representation for billions of years. This inference does not in itself show that truth cannot consist in a relation only among representations, because a proponent of the coherence theory could simply maintain that there were no representations and hence no true representations until intelligent beings evolved. But if there is a world independent of representation of it, as historical evidence suggests, then the aim of representation should be to describe the world, not just to relate to other representations. My argument does not refute the coherence theory, but shows that it implausibly gives minds too large a place in constituting truth.

Hence truth must consist of some sort of relation between the representations that occur in human minds or artifacts and the world. Truth is not a purely mental matter,

because our best evidence suggests that minds and their representations have not been around all that long. The advocate of the coherence theory of truth could desperately contend that truth is coherence in the mind of God, but this supposes that the most coherent view includes belief in the existence of God, a supposition that I have challenged (Thagard, 2000, ch. 4). There are of course intensely skeptical challenges that can be made to this use of scientific evidence, but I will hold off addressing these until section 10.

A different way of trying to protect the coherence theory of truth against my argument would be to say that the bearers of truth are not mental representations but eternal abstract entities. On this view, propositions are not mental structures, but Platonic objects constituting the meanings of sentences irrespective of whether there are any sentences. The problem with this Platonic reply is that we have no evidence that such abstract entities exist. In contrast, there is ample evidence from contemporary psychology and neuroscience that people employ mental representations, which therefore qualify as potential bearers of truth. Of course, the fact that, as far as we know, there were no mental representations 10 billion years ago does not undermine the correspondence theory of truth, because we can consider the fit, or lack of fit, between current representations and the state of reality at that time. The key point against the coherence theory of truth is that coherence with currently available evidence supports the view that reality is independent of representation of it.

At the other extreme from the coherence theory of truth, there is the view that coherence is simply irrelevant to truth. In epistemology, coherentist theories of knowledge are contrasted with foundational theories, which contend that knowledge is

based, not on a group of representations fitting with each other, but on a ground of indubitable truths. Rationalist foundationalists maintain that this ground consists of a set of truths known a priori, whereas empiricist foundationalists maintain that the ground is truths arising from sense experience. Unfortunately, both kinds of foundationalism have been dramatically unsuccessful in establishing a solid basis for knowledge. If there are any a priori truths, they are relatively trivial such as Hilary Putnam's (1983) principle that not every statement is both true and false. No one has succeeded in constructing a priori principles that receive general agreement and enable the derivation of substantial knowledge.

Similarly, the empiricist project of deriving knowledge from sense experience foundered because of the non-certain nature of sense experience and the non-derivability of scientific knowledge from experience alone. Our greatest epistemic achievements are scientific theories such as relativity theory, quantum theory, the atomic theory of matter, evolution by natural selection, genetics, and the germ theory of disease. None of these reduces to rationalist or empiricist foundations, so some kind of coherence theory of knowledge must be on the right track. (I am assuming that contextualism is a variety of of coherentism, but see Henderson, 1994.) Rejection of a connection between coherence and truth is therefore tantamount to adopting general skepticism about the attainability of scientific knowledge. In section 4, I will discuss skepticism arising from doubts about whether science really does attain truth.

I must also mention another prominent approach to relating coherence and truth that uses probability theory (e.g. Olsson, 2002; Shogenji, 1999). On this view, it should be possible to establish a connection between coherence and truth by means of an

intermediary connection between coherence and probability. If we could show that propositions with greater coherence have higher probability, then we could judge that they are more likely to be true. This is a laudable project, but I see three insurmountable problems with it: interpretation, realization, and implementation.

The interpretation problem for probabilistic epistemology is the need to choose what meaning to assign to probabilities, which may be taken either as frequencies or degrees of belief. The frequency interpretation of probability clearly applies well to scientific areas where data have been collected, but it does not apply at all to scientific theories. The probability of drawing a spade from a deck of cards is .25, meaning that in the long run the ratio of spades to cards drawn will approximate .25. But there are no comparable ratios applicable to scientific theories. On the other hand, viewing probabilities as degrees of belief is complicated by substantial evidence that human thinking does not naturally conform to Bayesian standards (Kahneman, Slovic, Tversky, 1982; Gilovich, Griffin, and Kahneman, 2002). Normatively, one might insist that it should, but there would still be the gap between subjective degrees of belief and objective truth. Using probability as the connection between coherence and truth presupposes that there are links between (1) coherence and probability and (2) probability and truth. The frequentist interpretation of probability fails in making the first link, whereas the degree-of-belief interpretation fails in making the second link.

By the realization problem I mean the difficulty of analyzing coherence in such a way that these links can be made. Ingenious analyses of coherence in terms of probability have been made by Olsson (2002) and Shogenji (1999); but on their own terms they have been unsuccessful in connecting coherence and truth, independent of the

problem of interpreting probability. A third problem, less commonly discussed in the philosophical literature, concerns the difficulty of implementing probabilistic reasoning computationally. Splendid computational tools have been developed for making inferences with probabilities (e.g. Pearl, 1988). But applying them to realistic cases of causal inference such as those involved in scientific or legal reasoning requires the concoction of large numbers of conditional probabilities that no reasoner would have available. For further discussion of the interpretation and implementation problems in the context of legal inference, see Thagard (2004).

Hence probability will not provide the needed connection between coherence and truth. Adherents of probabilistic epistemology would probably react by saying: so much the worse for coherence. But the intractability of the interpretation and implementation problems suggests a different response: so much the worse for probability, whose epistemological use is limited to areas like statistical inference where the frequency interpretation applies. In the rest of this paper, I will pursue a non-probabilistic approach to the connection between coherence and truth.

3. Explanatory Coherence

This pursuit requires much greater specification of what coherence is, and for that purpose I will employ my theory of explanatory coherence (Thagard, 1989, 1992, 2000). The theory consists of the following principles:

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; and (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.

These principles do not fully specify how to determine coherence-based acceptance, but algorithms are available that can compute acceptance and rejection of propositions on the basis of coherence relations. The most psychologically natural algorithms use artificial neural networks that represent propositions by artificial neurons or *units* and represent coherence and incoherence relations by excitatory and inhibitory links between the units that represent the propositions. Acceptance or rejection of a proposition is represented by the degree of activation of the unit. The program ECHO spreads activation among all units in a network until some units are activated and others are inactivated, in a way that maximizes the coherence of all the propositions represented

by the units. ECHO has been applied to many cases of scientific and legal reasoning, without the implementation and interpretation problems that afflict probabilistic models of causal reasoning. The theory of explanatory coherence depends on the notion of explanation, which section 7 discusses in terms of causal mechanisms.

The question now arises: do we have any reason to believe that a set of hypotheses that are accepted because they maximize explanatory coherence are at least approximately true? In ordinary life, counterexamples abound. For example, the theological and political beliefs of Osama bin Laden constitute a highly coherent set for him and his followers, but we would not want to acknowledge the truth of many of them. All of us have had the experience of making an inference to the best explanation about the behaviour of a friend or the breakdown of a piece of machinery, only to learn that our inference was erroneous. Such everyday cases are often deficient, however, in considering the full range of evidence and alternative hypotheses, so perhaps if explanatory coherence had been assessed properly the erroneous inference might have been avoided. But the history of science contains many cases where theories high in explanatory coherence have turned out to be false.

4. The Pessimistic Induction

Newton-Smith (1981, p. 14) named as the *pessimistic induction* the inference that any scientific theory will eventually be discovered to be false. Laudan (1981) compiled a long list of theories from the history of science that support this induction, including:

- the crystalline spheres of ancient and medieval astronomy
- the humoral theory of medicine
- catastrophist geology

- the phlogiston theory of chemistry
- the caloric theory of heat
- the vital force theory of physiology
- the aether theories of electromagnetism and optics
- theories of spontaneous generation.

The phlogiston theory, for example, had very substantial explanatory coherence, providing explanations of phenomena such as combustion and rusting that dominated chemistry for most of the eighteenth century. It was clearly the best explanation of the evidence until it was supplanted by Lavoisier's oxygen theory in the 1780s (Thagard, 1992, chs. 3-4). The pessimistic induction does not require that we know about the falsehood of previous theories by virtue of the truth of the theories that replaced them, which would make the induction incoherent. All it requires is noticing that many theories accepted as true were later rejected as false.

The pessimistic induction suggests that, since a great many theories in the history of science have turned out to be false, we should expect our current theories to turn out to be false as well. There may not be strong alternatives now for our most coherent theories such as relativity and evolution, but we should expect that eventually they will be superseded theories with more explanatory coherence. Hence we should not associate the maximization of explanatory coherence with truth. The history of science thus suggests that coherence, even explanatory coherence along the lines I have suggested, is a poor guide to truth.

Various responses are available to the pessimistic induction (Psillos, 1999). It is noteworthy that Laudan's examples are all from before the 20th century, and one could

argue that recent science has been more successful in achieving true theories. After all, the personal and material resources of science have increased steadily over the past 100 years. However, the temporal induction that recent theories will turn out to be true seems rather shaky, because there just might not have been enough time for superior theories to come along and demonstrate the weaknesses of current theories. Smolin (2001) suggests that problems in making quantum theory and relativity theory compatible with each other may lead to the replacement of both by a quantum theory of gravity. We need a more epistemologically satisfying induction that can tell us when we can take a coherent theory to be true.

5. Whewell's Overoptimistic Induction

William Whewell, the great nineteenth-century historian and philosopher of science, identified a feature of theories that he thought identified ones that are true. He used the term *consilience of inductions* to describe the support for a hypothesis that comes when it enables us to explain cases of a *kind different* from those which were contemplated in its formation (Whewell, 1968, p. 153). Whewell had a comprehensive knowledge of the history of science to the mid-nineteenth century, and he generalized as follows: “No example can be pointed out, in the whole history of science, so far as I am aware, in which this Consilience of Inductions has given testimony in favour of an hypothesis afterwards discovered to be false” (pp. 154-155). Consilience requires a hypothesis to increase its explanatory coherence, not merely by explaining a new fact, but by explaining a kind of fact different from ones previously explained. If Whewell were right, then we would have the basis for an optimistic induction about the relation of

coherence and truth: Theories that become more coherent over time by explaining new kinds of facts turn out to be true.

Whewell's favorite examples of consilience were Newton's theory of universal gravitation and the undulatory (wave) theory of light. A century and a half after Whewell's generalization, these examples appear rather unfortunate. Newton's theory of gravitation was never able to explain the perihelion of Mercury, but Einstein showed in 1915 that his theory of general relativity yielded more accurate predictions of it than did Newton (Gondhalekar, 2001). Thus it appears that, contrary to Whewell's optimistic induction about a particular kind of coherence signaling truth, we have yet another instance of the pessimistic induction. Einstein rejected Newton's assumption that gravitational force is a function only of mass and distance, along with his assumptions of absolute space and time. Similarly, the wave theory of light has been superseded by the quantum theory that views light as consisting of particle-like photons that have wave-like properties.

An even more damning counterexample to Whewell's optimistic induction is the phlogiston theory, which was originally developed to explain combustion by Johann Becher, who called the principle of inflammability *terra pinguis*. Becher and Georg Stahl, who renamed this substance *phlogiston*, applied it also to explain calcination (rusting) and respiration. Since rusting and breathing appear very different from burning, extension of the phlogiston theory to calcination seems to constitute an instance of the consilience of inductions; but already by Whewell's time the falsity of the phlogiston theory had been recognized (Partington, 1961).

Thus Whewell was overoptimistic about the epistemic power of the consilience of inductions. Still, I think he was on the right track in looking for temporal properties of developing theories that might mark them as good candidates for truth. At least we can say that the theory of universal gravitation and the wave theory of light are not so totally false as the theories of crystalline spheres, phlogiston, and caloric turned out to be. Below I will try to identify a sense in which Newtonian gravitation is partly true, but first I want to discuss a truth-related mark of coherent theories that is more promising than consilience.

6. Deepening and the Cautiously Optimistic Induction

There are two main ways in which a hypothesis can increase its explanatory coherence over time. The first is to explain new facts, which I will call *broadening*. Whewell's consilience is a special kind of broadening in which the new facts explained are of a kind different from those already explained. The consilience of Wilson (1998) is an even more special kind of broadening involving the interlocking of causal explanations across disciplines.

The second way in which a hypothesis can increase its explanatory coherence is by being explained by another hypothesis, which I will call *deepening*. The process of a hypothesis being explained is most easily understood by legal examples in which questions of motive arise. The hypothesis that an accused is responsible for killing a victim gets its main explanatory coherence from its ability to explain a range of evidence, for example that the accused's fingerprints are on the murder weapon. This hypothesis can be broadened by finding new evidence that it explains, for example that one of the victim's hairs was found on the accused's clothes. Homicide detectives also want to

know about possible motives of the accused, in order to provide an explanation of why the accused committed the murder. For example, the case that O. J. Simpson killed his ex-wife was based mainly on the evidence that he did so, but also on his having the motive of jealousy for attacking her and her boyfriend. We then get a deeper explanation of the crime, because the motive explains the killing which explains the evidence. In the normal practice of the law, the explanation draws on folk rather than on scientific psychology, attributing the behavior of the accused to ordinary beliefs, desires, and emotions rather than on richer structures and processes for which there is experimental evidence.

Normally, the deeper hypothesis is not adduced merely for the purpose of explaining the basic one, but has independent evidence supporting it. Figure 1 shows the structure of a hypothesis A that has broadened by explaining pieces of evidence 2 and 3 in addition to evidence 1, and deepened by being explained by hypothesis B.

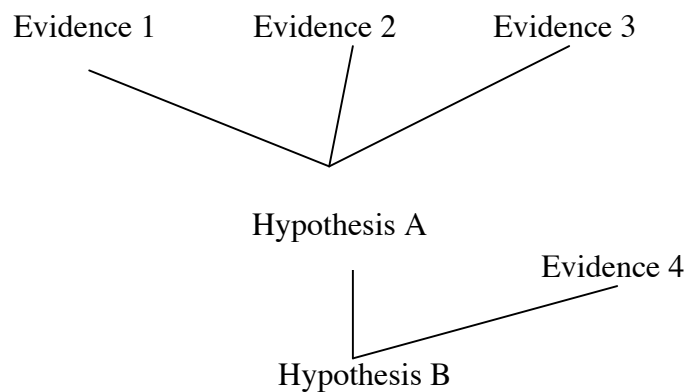


Figure 1. Hypothesis A is broadened by explaining new kinds of evidence, and deepened by being explained by hypothesis B that explains additional evidence.

In science, deepening occurs when an explanation provides an underlying causal basis for a causal hypothesis. For example, consider the germ theory of disease, which says that contagious diseases are caused by micro-organisms such as bacteria, viruses, and fungi. (To modern ears, this might sound like a tautology, but contagion was recognized long before the pathological effects of germs.) A particular instantiation of the germ theory identifies a specific disease that is caused by a specific microbe; for example, influenza is caused by viruses of the family *Orthomyxoviridae*. This theory has been deepened over the years by microbiological accounts that explain how viruses infect cells, replicate themselves, and disrupt the functions of cells and organs. For many viruses, biologists have identified all of their genes and the functions they perform. Thus we know not only that myxoviruses cause different types of influenza, but how they do so by their mechanisms of attachment, infection and reproduction. Some other examples of deepening in recent history of science include the use of microbiology to explain how genetic transmission works and the use of the quantum-mechanical theory of molecular bonding to explain how atoms combine into molecules, which explains molecular theories of chemical reactions.

Now let me venture my own version of an optimistic induction about the relation between coherence and truth, which I call the *deepening maxim*:

Explanatory coherence leads to truth when a theory is not only the best explanation of the evidence, but also broadens its evidence base over time *and* is deepened by explanations of why the theory works.

I am unaware of any broadened and deepened theory that turned out to be mostly false, but am aware that counterexamples may arise, so let me call this the *cautiously* optimistic

induction. Actually, we do not need a universal generalization here: it would be enough if we could show from a survey of the history of science that broadened and deepened theories rarely turn out to be false. It is remarkable that none of the theories that I discussed in connection with the pessimistic induction that theories turn out to be false were ever deepened. That is, no underlying mechanisms were identified for how entities such as phlogiston and caloric worked. The deepening maxim is a generalization not only about past theories but about future ones, predicting that future theories that are broadened and deepened will tend to be true.

My response to the pessimistic induction is very different from one recently criticized by Stanford (2003). He argues that scientific realism is undermined rather than supported by attempts to show that discarded theories such as phlogiston and caloric were at least partially successful with respect to the reference of central terms, core causal descriptions, partial truth, select preservation, and/or historical continuity. Instead of trying to defeat the pessimistic induction by arguing that the discarded theories are at least partly true, my strategy is to admit their falsehood and look for features that mark current theories as promising candidates to avoid joining phlogiston and caloric in the dustbin of history. The combination of broadening and deepening seems to be such a feature, but more needs to be said about how mechanistic explanations are deepened.

The importance of broadening and deepening is entailed by the theory of explanatory coherence discussed in section 3. It follows from principle E2, Explanation, that the more a hypothesis explains, the more coherent it is, other things being equal. Hence finding a new fact that is explained by a theory increases its explanatory coherence. Principle E2 also implies that a hypothesis explained by another hypothesis

is more coherent, other things being equal. However, the theory of explanatory coherence by itself is neutral about the nature of explanation, but fits well with the view that explanations are based on causal mechanisms.

7. Mechanisms and Explanation

To explicate the deepening maxim further, I need to say much more about how a theory can receive a deeper explanation. This requires an account of the nature of explanation, which can be provided by attention to the nature of mechanisms. It then becomes possible to characterize the nature of approximate truth in terms of the ontology of mechanisms.

In accord with much recent work in the philosophy of science, I hold that to explain a phenomenon is to describe a mechanism that produces it (Salmon, 1984; Bechtel and Richardson, 1993; Machamer, Darden, and Craver, 2000; Glennan, 1996; Thagard, 2003; Bechtel and Abrahamsen, forthcoming). A mechanism is a system of parts whose properties and relations produce regular changes in those properties and relations. For example, we explain how brains work by specifying their parts, which are neurons organized into neuronal groups and functional areas. Neurons have properties such as their electric potentials, and relations such as their links to other neurons via excitatory and inhibitory synapses. We can use mathematical equations and computer models to infer the behavior of groups of neurons from their properties and relations, including inputs from an external environment.

If explanation is mechanism-based, we can develop an account of deepening in terms of parts. A deeper explanation for an explanatory mechanism M1 is a more fundamental mechanism M2 that explains how and why M1 works. M1 consists of

parts, and M2 describes parts of those parts whose properties and relations change in ways that generate the changes in the properties and relations of the parts in M1. For example, neural mechanisms are deepened by noting that neurons consist of parts, proteins and other molecules that are organized into functional areas such as the nucleus, mitochondria, axons, dendrites, and synapses. Chemical reactions involving these proteins enable nerve cells to carry out their basic functions, including taking inputs from presynaptic neurons, building up electric charges, spiking, and sending signals to postsynaptic neurons. Thus neuroscience is deepened by molecular biochemistry which explains how neurons work. This is a special case of how cell biology is deepened by molecular biochemistry, as has progressively happened over the past half-century. Similarly, medical theories such as the germ theory of disease have been deepened by finding lower-level mechanisms for the operations of cells and their microbial invaders. As I mentioned in the last section, medicine knows enough about the parts of bacteria, viruses, and fungi to be able to explain how they invade and disrupt the normal function of bodily cells.

Deepening is also pervasive in modern chemistry and physics. We explain the chemical behavior of elements and compounds by the atoms that compose them. In turn, contrary to the original ancient and nineteenth-century atomic theories, the behavior of atoms can be explained by describing their parts, right down to the level of quarks and possibly superstrings. I am not assuming the traditional philosophical view of reductionism, according to which the deeper theory serves to deductively predict what goes on at the higher level: the complexity and sensitivity to chaos of higher and lower level systems may make such predictions impossible to achieve. But we gain much

understanding nevertheless by noting that the mechanisms at the upper level works as it does because of the operations of the parts at the lower level. The deepening maxim obviously does not apply to the most fundamental level in subatomic physics, but still has ample room for application in other areas of physics as well as chemistry, biology, and the social sciences. The instances claimed to support the pessimistic induction in section 4 are not at the fundamental level to which deepening cannot apply.

Thus in important areas of medicine, biology, chemistry, and physics, we commonly get deepening by theoretical mechanisms that show how the parts, properties, relations, and changes at the higher level decompose into parts at the lower level. Evidence for the lower level mechanism is not just that it explains the higher level one. For example, there is ample molecular and chemical evidence for the structure and operations of neurons and the operations of microbes, so molecular biochemistry provides support for the acceptability of the neuronal theory of brain operation and for the germ theory of disease. The deepening maxim can then be specified as the induction that theories can be judged to be true if they have been deepened by having the mechanisms they describe decomposed into more fundamental mechanisms for which there is independent evidence. As we have seen, inductive support for the deepening maxim includes the germ theory of disease, the neuronal theory of brains, molecular cell biology, molecular genetics, and the atomic theory of matter. Attention to mechanistic explanation serves to spell out and support my cautiously optimistic induction about the connection between coherence and truth.

8. Approximate Truth

However, there are some cases from the history of science that may constitute challenges to the deepening maxim and require weakening it to conjecture a connection only between coherence and approximate rather than absolute truth. Consider, for example, the atomic theory of matter. I used this as an example of deepening, because we can now explain how atoms undergo changes such as forming molecules by using quantum mechanical theories about their parts. But I also noted that this deepening required abandonment of the previously definitional truth that atoms do not have parts. So strictly speaking, the atomic theory was not deepened, but rendered false. It would be a mistake, however, to treat a theory as merely a conjunction of hypotheses that is false if any one of them is false. Instead, we should spell out what the theory claims about mechanisms consisting of parts, properties, relations, and resulting changes, and identify how many of these claims turn out to be wrong. We can then maintain that the atomic theory of matter has survived because most of its claims about the constituents of things are still taken to be true: elements and compounds have properties, relations, and changes that result from the atoms of which they consist.

Newtonian mechanics is another difficult case. I used it to challenge as overoptimistic Whewell's induction about consilience. But it is also a possible challenge to my deepening maxim, because it might be argued that general relativity deepens Newtonian theory by providing an explanation in terms of the curvature of space-time of how Newton's force of gravity works. Then Newton's theory of gravitation seems like a counterexample to the deepening maxim, in that it has been deepened but is acknowledged to be false by virtue of its failed predictions and rejected assumptions such as that mass is independent of energy.

However, to say that general relativity totally replaced Newtonian gravitation would be as much of a mistake as saying that it fully incorporates it (see Thagard, 1992, pp. 214-215 for the details). Newton's three laws and the principle of gravitation are good approximations to what general relativity predicts as long as velocities and masses are small. Although Newtonian mechanics does not accurately predict the perihelion of Mercury, its degree of inaccuracy was only 8%: 43 arcseconds per century (Gondhalekar, 2001, p. 162). So it is reasonable to maintain that Newtonian mechanics is approximately true in the sense that its major claims are quantitatively close to those supported by evidence and the theory that replaced it. I take a theory to be approximately true if it is *partly* true, i.e. if *most* of its claims are *nearly* true in achieving quantitative closeness to accepted values. Assessment of approximate truth does not simply involve counting sentences, but needs to qualitatively consider the central mechanistic claims that the theory makes about parts, properties, relations, and resulting changes. For further discussion of approximate truth, also known as verisimilitude and truthlikeness, see Psillos (1999), Kuipers (2000), and Thagard (2000).

My cautiously optimistic induction is thus cautious in two respects. First, it allows for the possibility that a major instance of a deepened theory could turn out to be false. I do not expect fields such as molecular medicine, genetics, and atomic theory to be radically overthrown, but it could happen. Second, it allows for the possibility, which seems to have happened in both the atomic theory of matter and Newtonian mechanics, that deepening by virtue of a more fundamental mechanism may lead to some revisions in the original theory, with recognition that it is only approximately (partly and nearly) true. Accordingly, here is my final version of the deepening maxim: If a theory

not only maximizes explanatory coherence, but also broadens its evidence base over time *and* is deepened by explanations of why the theory's proposed mechanism works, then we can reasonably conclude that the theory is at least approximately true. This induction is the strongest relation available between coherence and truth.

If a theory is broadened and deepened, is it still the same theory? As in the discussion of approximate truth, it is useful to think of a theory not just as a set of sentences but rather as a representation of parts, properties, relations, and changes. Broadening a theory by finding a new explanatory application of it and deepening it by identifying an underlying mechanism clearly do not generate a new theory, as long as the original hypotheses about parts, properties, relations, and changes remain substantially the same.

9. Deepening the Deepening Maxim

The deepening maxim gains credibility from the numerous theories in the history of science, such as the germ theory of disease and the atomic theory of matter, that have both undergone deepening and avoided the dustbin of rejected theories. But it would gain increased credibility if we could say *why* it is true that deepened theories tend to hold up well to empirical investigation and therefore appear to be approximately true. Deepening is not a necessary condition of the acceptability of a scientific or philosophical claim, because knowledge about underlying mechanisms may simply not be available at a given time. The germ theory of disease, the theory of evolution by natural selection, and atomic theory were all credible before microbiology, genetics, and subatomic physics were developed. Nevertheless, they gained additional credibility when underlying mechanisms became understood.

Hence it is useful to ask why deepened theories tend to avoid becoming instances that support the pessimistic induction. One superficial explanation might be that deepened theories tend to survive because scientists simply prefer deepened theories as part of their overall strategy of maximizing explanatory coherence. Deepened theories survive not because they are good candidates for approximate truth, but merely because they are popular with scientists. The flaw in this explanation is that it ought also to apply to broad theories: broad theories survive just because they are what scientists like, not because they have anything to do with truth. But we have seen abundant examples of broad theories that were superseded by broader theories, as oxygen superseded phlogiston and thermodynamics replaced caloric. Even theories that underwent broadening – Whewell’s examples of Newtonian mechanics and the wave theory of light – have given way to successors.

Following a suggestion of Peter Railton’s (personal communication), I think that the most plausible explanation of why deepened theories survive and thrive is that they are at least approximately true. The success of theories for whom underlying mechanisms are found is not the result of scientific fashion, but rather the result of a strategy that fits well with the structure of the world that science investigates. Different kinds of experiments and instruments enable science to study the structure of the world at different levels, and these levels fit together naturally. For example, when a health researcher can transfer illness from one animal to another, the tools of bacteriology and virology explain why this is so. In turn, the tools of molecular biology enable an explanation of how viruses and bacteria infect cells by means of molecular signals. In turn, the tools of physical chemistry enable an explanation of how atoms form molecules

and how chemical reactions occur. Here the deepening strategy works because the world is in fact organized in terms of parts, from organisms down to subatomic particles, and layers of mechanisms, from viral infection down to chemical bonding. Hence the best explanation of the great success of deepened theories is that they are finding out about the world at multiple levels. Scientific realism should not consider individual theories in isolation from each other, but should notice how well the most successful of them manage to nest vertically with each other as well as to fit with experimental observations. This nesting, accomplished by instruments and experiments and theorizing that operate at multiple interacting levels, is the key to a deep account of the relation between coherence, deepening, and truth.

The argument in this section is a special case of the general abductive argument for scientific realism, which has the form:

- (1) Science is successful with respect to predictions, technological applications, cumulation of knowledge, and agreement among practitioners.
- (2) The best explanation of this success is that scientific theories are at least approximately true.
- (3) Therefore, scientific theories are at least approximately true.

This is not the place to discuss or defend the general argument (see Thagard, 1988, ch. 8; Kitcher, 2002). I am presenting a more specific argument:

- (4) Deepening maxim: Theories that have undergone deepening by lower-level mechanisms have survived empirical investigation.

(5) The best explanation of this survival is that scientific theories about layers of mechanisms are at least approximately true.

(6) Therefore, scientific theories about layers of mechanisms are at least approximately true.

Notice that (6) follows deductively from (3), so that the general support for scientific realism supports scientific realism about mechanisms which then supports the deepening maxim. The support is mutual, as is generally the case in explanatory coherence: When a hypothesis explains evidence or another hypothesis, both the explainer and the explained support one another. In the current context, the crucial point is that the deepening maxim has been deepened by a hypothesis about the approximate truth of theories about layers of mechanisms.

Levin (1984) challenged scientific realists to give *mechanistic* explanations of how truth produces success, which would require at least a sketch of the mechanisms by which the truth of theories about layers of mechanisms produces the success described by the deepening maxim. This task is daunting, but should be doable for particular cases such as my medical example relating infectious diseases to underlying mechanisms. The relevant mechanisms include use of instruments that detect features of the world at different levels of detail, ranging from stethoscopes to MRI machines to optical and electron microscopes. Defending the reliability of these instruments provides further support for scientific realism, in accord with the “Galilean strategy” of Kitcher (2001). Additional relevant mechanisms include physical and social processes by which experiments are carried out, and cognitive mechanisms by which human minds collect and interpret data. Thus filling out the explanation in (5) that links approximate truth

with deepening is very complex, but there is reason to believe that it could be done mechanistically.

10. Conclusion

I have argued that many of the standard claims about the relation of coherence to truth – that coherence is truth, that coherence is irrelevant to truth, and that coherence leads to probability which leads to truth - are implausible. Instead, I have built upon my theory of explanatory coherence and the mechanism-based theory of explanation deepening to produce a cautiously optimistic induction about when coherence usually leads to truth in natural science. This induction constitutes a response to doubts whether coherence connects to truth. (Millgram, 2000, raises the question of whether *approximation* to coherence of the sort performed by available algorithms leads to truth; for a response, see Thagard, unpublished.)

To conclude, I will discuss the implications of this view of the relation between coherence and truth for social science, everyday life, and philosophical deliberation about the nature of knowledge. Social science has not witnessed the extent of deepening found in the natural sciences, but there are some good prospects. Cognitive psychology is now heavily enmeshed with neuroscience, and cognitive theories, which were previously couched only in terms of representations and computations, are increasingly being fleshed out in terms of neural structures and processes. The neurological turn in cognitive psychology, and to a lesser extent in other areas of psychology such as social, developmental, and clinical, opens the possibility of the field being deepened by neuroscience, just as neuroscience is being progressively deepened by neurochemistry. Another promising trend is that economics, which formerly dealt with highly idealized

models of human rationality that had little to do with human psychology, is increasingly tied with cognitive psychology and neuroscience, through the development of the field of behavioral economics. I am not raising the prospect of reducing economics to psychology and psychology to neuroscience, but rather pointing to the salutary trend of economics enriching its theories by information about psychological mechanisms, and psychology enriching its understanding by information about neurological mechanisms, which in turn are deepened by biochemical mechanisms. Hence there is a reasonable prospect that my cautiously optimistic induction may eventually apply to social science as well as natural science.

What about everyday life? Most epistemologists have worried about people's ordinary knowledge, such as what we are purported to know about the external world and other people. Although the deepening maxim does not apply to most beliefs of most people, this does not mean that they are grossly unwarranted. The deepening maxim specifies conditions under which coherence very reliably leads to truth, but it does not imply that there is no truth without deepening. Broad explanatory theories in everyday life may well turn out to be often true, for example our common, everyday beliefs about physical objects and the behavior of other people.

More importantly, if we step outside everyday beliefs and detect their origins in psychophysical processes, we can deepen everyday explanations considerably. For example, beliefs about physical objects that arise from sense perception can be understood more deeply by going beyond ordinary people's knowledge about seeing, through appreciation of the physical and neurological processes that connect objects and our perception of them by photons of light, retinal stimulation, and neural processing.

Physics and neuroscience then provide mechanisms explaining why people have experiences and beliefs about the external world, thereby deepening their knowledge about it. Similarly, scientific psychology and neuroscience provide a basis for explaining why folk psychology is sometimes right about why people think and behave the ways they do, and sometimes wrong.

Whereas common sense perception and cognition are open to deepening by physical and neurological mechanisms, skeptical theories are not. Skeptics have suggested that our experiences may arise, not from veridical psychological processes, but by the machinations of a deceptive god, evil genius, or matrix of brains in vats. These alternative theories are inherently shallow, in that there are no evidence-supported mechanisms that explain how we could be so systematically deceived. To consider only the most recent example, the hypothesis from the *Matrix* movies, that human experiences arise by illusory inputs from intelligent machines, fails to take into account the physical and computational implausibility of generating the complexity and rapidity of such experiences. It takes years to produce an animated movie such as *Finding Nemo*, so there is no plausible technology that could produce the multiple streams of interconnected experiences portrayed in the *Matrix*. Thus common sense epistemology is open to deepening by reasonable extensions of current physics, psychology, and neuroscience, whereas skeptical epistemologies float flimsily in the air. Hence developments in the cognitive sciences should eventually have the result that the everyday hypothesis that our perceptions arise largely as the result of interactions with an external world will fall within the scope of the cautiously optimistic induction. Then naturalistic epistemology

will mesh with the philosophy of science to provide a deep justification of everyday as well as scientific knowledge.

Acknowledgements. For helpful suggestions, I am grateful to Josh Brown, Dave DeVidi, Eric Lormand, Elijah Millgram, Eric Olsson, Peter Railton, and anonymous referees.

References

- Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity*. Princeton: Princeton University Press.
- Bechtel, W., & Abrahamsen, A. A. (forthcoming). Explanation: A mechanistic alternative. *Studies in History and Philosophy of Science*.
- Blanshard, B. (1939). *The nature of thought, vol. 2*. London: George Allen & Unwin.
- BonJour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.
- Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49-71.
- Goldman, A. (1999). *Knowledge in a social world*. Oxford: Oxford University Press.
- Gondhalekar, P. (2001). *The grip of gravity: The quest to understand the laws of motion and gravitation*. Cambridge: Cambridge University Press.
- Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press/Bradford Books.
- Henderson, D. K. (1994). Epistemic competence and contextualist epistemology: Why contextualism is not just the poor man's coherentism. *Journal of Philosophy*, 91, 627-649.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

- Kitcher, P. (2001). Real realism: The Galilean strategy. *Philosophical Review*, 110, 151-197.
- Kitcher, P. (2002). On the explanatory role of correspondence truth. *Philosophy and Phenomenological Research*, 64, 346-364.
- Kuipers, T. (2000). *From instrumentalism to constructive realism*. Dordrecht: Kluwer.
- Laudan, L. (1981). A confutation of convergent realism. *Philosophy of Science*, 48, 19-49.
- Lehrer, K. (1990). *Theory of knowledge*. Boulder: Westview.
- Levin, M. (1984). What kind of explanation is truth? In J. Leplin (Ed.), *Scientific realism* (pp. 124-139). Berkeley: University of California Press.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1-25.
- Millgram, E. (2000). Coherence: The price of the ticket. *Journal of Philosophy*, 97(82-93).
- Newton-Smith, W. H. (1981). *The rationality of science*. London: Routledge & Kegan Paul.
- Olsson, E. J. (2002). What is the problem of coherence and truth? *Journal of Philosophy*, 99, 246-272.
- Partington, J. (1961). *A history of chemistry*. London: Macmillan.
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems*. San Mateo: Morgan Kaufman.
- Psillos, S. (1999). *Scientific realism: How science tracks the truth*. London: Routledge.

- Putnam, H. (1983). There is at least one *a priori* truth. In H. Putnam (Ed.), *Realism and reason: Philosophical papers, volume 3* (pp. 98-114). Cambridge: Cambridge University Press.
- Rescher, N. (1973). *The coherence theory of truth*. Oxford: Clarendon Press.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- Shogenji, T. (1999). Is coherence truth-conducive? *Analysis*, 61, 338-345.
- Smolin, L. (2001). *Three roads to quantum gravity*. New York: Basic Books.
- Stanford, K. S. (2003). Pyrrhic victories for scientific realism. *Journal of Philosophy*, 100, 553-573.
- Thagard, P. (1989). Explanatory coherence. *Behavioral and Brain Sciences*, 12, 435-467.
- Thagard, P. (1992). *Conceptual revolutions*. Princeton: Princeton University Press.
- Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.
- Thagard, P. (2003). Pathways to biomedical discovery. *Philosophy of Science*, 70, 235-254.
- Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, 18, 231-249.
- Thagard, P. (unpublished). *Coherence: The price is right*. Retrieved Aug. 31, 2006, from <http://cogsci.uwaterloo.ca/Articles/Pages/coh.price.html>
- Whewell, W. (1968). *William Whewell's theory of scientific method*. Pittsburgh: University of Pittsburgh Press.
- Wilson, E. O. (1998). *Consilience: The unity of knowledge*. New York: Vantage.

Young, J. O. (2001). *The coherence theory of truth*. Retrieved August 6, 2005, from
<http://plato.stanford.edu/entries/truth-coherence/>