Topics in Cognitive Science 1 (2009) 237–254

ISSN: 1756-8757 print / 1756-8765 online
DOI: 10.1111/j.1756-8765.2009.01016.x

Why Cognitive Science Needs Philosophy and Vice Versa

Paul Thagard

Department of Philosophy, University of Waterloo

Received 27 January 2009; received in revised form 11 February 2009; accepted 11 February 2009

Abstract

Contrary to common views that philosophy is extraneous to cognitive science, this paper argues that philosophy has a crucial role to play in cognitive science with respect to generality and normativity. General questions include the nature of theories and explanations, the role of computer simulation in cognitive theorizing, and the relations among the different fields of cognitive science. Normative questions include whether human thinking should be Bayesian, whether decision making should maximize expected utility, and how norms should be established. These kinds of general and normative questions make philosophical reflection an important part of progress in cognitive science. Philosophy operates best, however, not with a priori reasoning or conceptual analysis, but rather with empirically informed reflection on a wide range of findings in cognitive science.

Keywords: Cognitive science; Philosophy; Generality; Normativity; Theories; Explanations; Computer simulation; Bayesian inference; Decision making

1. Introduction

A prominent cognitive scientist once told me that philosophy is to cognitive science what tin cans tied to a car are to a wedding. An equally negative analogy produced by a psychologist is that philosophy is to science as alcohol is to sex, which I take to mean (in Shakespeare's words from *Macbeth*) that it provokes the desire but takes away the performance. It has also been claimed that philosophy is to science as pornography is to sex. Richard Feynman is supposed to have said that scientists are explorers, but philosophers are tourists, and that philosophy of science is about as useful to scientists as ornithology is to birds.

Correspondence should be sent to Paul Thagard, Philosophy Department, University of Waterloo, Waterloo, ON N2L 3G1, Canada. E-mail: pthagard@uwaterloo.ca

To all these ugly comparisons, I prefer a different view adapted from Santayana's famous remark about history: Those who ignore philosophy are condemned to repeat it. To amplify, I adapt a remark of Keynes about economic theory: Those who believe themselves to be exempt from philosophical influence are usually the slaves of some defunct philosopher.

My aim in this paper is to show that philosophy is essential to the interdisciplinary study of mind, but not for the reasons that many philosophers assume. Philosophy does not provide foundations for cognitive science and is incapable of generating the a priori truths that many philosophers have sought. Philosophy is not the queen of the sciences. Nor does philosophy have a special role in clearing up conceptual confusions about the study of mind, as this alleged role misunderstands the nature of concepts.

Rather, philosophy has two major contributions to make to cognitive science: generality and normativity. By generality I mean that philosophical reflections attempt to answer questions that are broader than those usually pursued by researchers in particular disciplines such as psychology, neuroscience, linguistics, anthropology, and artificial intelligence. Philosophical generality is especially crucial for an interdisciplinary field such as cognitive science, in that it can attempt to address questions that cross multiple areas of investigation, thereby helping to unify what otherwise appear to be diverse approaches to understanding mind and intelligence. I will provide examples of how philosophy can aim to answer general questions about the investigation of mind that inevitably arise in the most ambitious scientific investigations of thought.

By normativity I mean that philosophy is concerned not only with how things are but also with how they should be. Philosophical theories of knowledge and morality need to go beyond descriptive theories of how people think and act by also developing normative (prescriptive) theories of how people ought to think and act. I will argue that normative issues are unavoidable in cognitive science, requiring the participation of philosophers who possess some of the theoretical tools needed to address them. Many philosophers have thought that, in order to pursue this normative function, philosophy must distance itself from empirical matters, but I will provide examples where the investigation of descriptive and normative issues go hand in hand.

In addition to generality and normativity, philosophy can play other subsidiary roles in cognitive science that I merely mention. Sometimes philosophical ideas can be useful in stimulating scientific investigations, for example, when some of Wittgenstein's ideas about language inspired important 1970s research about the prototypical nature of concepts, and when Daniel Dennett's views of intentional action triggered a flourishing research tradition in developmental psychology concerned with children's judgments about false beliefs. Philosophy can be also useful to cognitive science in providing defenses against philosophical arguments challenging the core assumptions of cognitive science concerning representation and computation. In this way, philosophy can provide self-defense methods for cognitive scientists against philosophers critical of the whole field. I conclude this essay by showing how philosophy depends on cognitive science in its attempt to develop general and normative theories about knowledge, reality, morality, and meaning.

2. Generality

Whenever science operates at the edge of what is known, it runs into general issues about the nature of knowledge and reality. Mundane science can operate without much concern for methodological and ontological issues, but frontier science cannot avoid them. For example, innovative research in theoretical and experimental physics inevitably encounters fundamental problems about the nature of space and time, as well as methodological questions concerning how scientific investigation should proceed. Cognitive science has made substantial progress in investigating phenomena such as perception, memory, learning, problem solving, and language use, but clearly it is still a frontier enterprise.

Here are some central questions that arise in cutting-edge research in cognitive science: What is an explanation? What is a theory? How can competing theories be evaluated? What is the relation between different fields of cognitive science such as psychology and neuroscience? What is the role of computer modeling in cognitive science? I will not attempt to answer these questions here, but I will provide some pointers to philosophical work that addresses them in ways that I think are useful for the self-understanding of cognitive science research, encouraging progress rather than hindering it. This list of questions is not intended to be complete, but it provides a sample of the sort of philosophical issues encountered by cognitive science at its most interesting.

2.1. What are theories and explanations in cognitive science?

Cognitive science operates with a multitude of conferences, societies, journals, research groups, and educational programs, but it rarely stops to ask what is the point of all this activity. In my view, the main aims of interdisciplinary research on mind and intelligence are to understand how the human mind works and to use this understanding to develop ways of making humans and machines more intelligent. If you disagree with this opinion, then we need to have a very general and hence philosophical discussion about what cognitive science is for, a question that cannot be answered just by the experimental and theoretical methods of cognitive science.

Cognitive science provides understanding by giving accounts of the nature of key phenomena such as inference. Such accounts are provided by theories that can be used to explain the results of experiments. Here are some possible answers to the question of what constitutes a scientific theory:

1. A cognitive theory is a set of mathematical formulae used to make predictions about behavior.
2. A theory is a computer program that simulates thinking.
3. A cognitive theory is a description of mechanisms that explain observed mental phenomena.

I prefer the third view of theories, because it fits well with the most successful practices in psychology and neuroscience, as well as related areas in biology and medicine (see, for

example, Bechtel, 2008; Bechtel & Richardson, 1993; Craver, 2007; Darden, 2006; Thagard, 1999, 2005, 2006). But the first two options have also been assumed by leading contributors to cognitive science, so there is a philosophical issue here that deserves to be debated by anyone reflective about how the mind is to be understood.

In physics, it is often claimed that the primary function of a theory is to generate predictions, but explanations are at least as important. There is a long history of philosophical discussion of the relation between prediction and explanation (e.g., Hempel, 1965). The making of precise predictions is much more difficult in the biological than in the physical sciences, but explanations are just as important. Here are some candidate answers to the question of what is an explanation:

1. An explanation is an answer to a question about why something happened.
2. An explanation is a deductive derivation of a description of a phenomenon from a set of principles.
3. An explanation is a description of how the operation of a mechanism produces a phenomenon.

Consonant with my preferred view of theories, I see explanations as primarily mechanistic, but this view is highly controversial in the philosophy of science, and I expect that many cognitive scientists would find it puzzling or possibly odious.

In reply, I would argue that the most successful theoretical explanations in cognitive science, for example, using rule-based and connectionist ideas, have been mechanistic in the sense elucidated by philosophers of science. A mechanism is a system of parts whose interactions produce regular changes. Rule-based systems such as GPS (Newell & Simon, 1972) and ACT (Anderson, 2007) are clearly mechanistic in this sense, as are neural networks models such as PDP models (Rumelhart & McClelland, 1986) and more recent models closer to actual brain mechanisms (e.g., Eliasmith & Anderson, 2003). Then the primary difference between conflicting theories is the postulation of different kinds of parts and interactions as responsible for the psychological phenomena that everyone wants to explain.

How can cognitive science adjudicate between competing theories about the mechanisms that make mind works? There are various philosophical answers to this question, ranging from hypothetico-deductive (Popper, 1959) to Bayesian (Sober, 2008) to explanatory coherence (Thagard, 1992). This is not the place to defend my own favorite answer, which is based on a cognitive model that has had many philosophical and psychological applications. My point is that the issue of theory evaluation supports my Santayana- and Keynes-inspired remarks in the first paragraph. When cognitive scientists do not have explicit views about the methodological issues concerning theories, explanations, and evaluations, it is not because they do not have any views, just that the views they hold are usually implicit and unreflective. Methodological issues inevitably arise in key disputes in cognitive science, for example, concerning mental imagery and the value of abstract Bayesian models that assume that optimality is a property of human information processing. Such issues need to be confronted by scientists with the assistance of philosophy, not buried as if the answers to the philosophical questions were obvious.

Explanations in cognitive science often help themselves to the concept of causality without addressing longstanding philosophical issues about its nature. Is causality a matter of constant conjunction, mental schemas, probability, special powers, manipulability, energy transfer, or nothing at all? My own view is that the basic human concept of causality is a complex multimodal, neural representation of changes that includes both sensorimotor encodings of manipulability and verbal encodings of regularity (Thagard & Litt, 2008). Regardless of whether this view is correct, cognitive science needs to blend its frequent use of causality in explanations of human thought and behavior with philosophical reflection about what causality amounts to.

Philosophers do not have some special, a priori ability to settle questions about the nature of theories and explanations. But they have (or should have) awareness of a wider range of answers to these questions, as well as familiarity with the main answers that have been proposed in the past. Moreover, philosophy has been accustomed to ask such questions in full generality, so that they cover other sciences such as physics and biology as well as the various disciplines that make up cognitive science.

2.2. What is the role of computer modeling in cognitive science?

The use of computer models has been an important feature of cognitive science since Newell, Shaw, and Simon (1958) produced the first computational account of human problem solving. The organizational beginnings of cognitive science in the late 1970s, heralded by formation of the journal Cognitive Science and the Cognitive Science Society, explicitly looked for research that combined psychology and artificial intelligence. Since then, computational models have flourished in many interdisciplinary branches of cognitive science, including computational neuroscience, computational linguistics, computational organization science, and even computational philosophy. But what are these computer models actually contributing to cognitive science?

It is sometimes said that a computer program can be a cognitive theory, but I think this is a mistake, because programs are always full of minor details tied to the particular kind of programming language they are written in. For example, a program written in the common AI language LISP will have lists as its most common data structure, but I know of no LISP modeler who claims psychological significance for this particular form of representation. Rather, it is crucial to distinguish between theories, models, and programs (Thagard, 2005).

In keeping with my preferred account in the previous section, I think a theory is a description of an explanatory mechanism, consisting of parts and their interactions producing regular changes. In a psychological theory, the parts are mental representations such as concepts, rules, and images; and the interactions are processes such as spreading activation and rule matching and firing. In a neural theory, the parts are neurons and neural groups, and the interactions include excitation, inhibition, and learning by weight changes. In an organizational theory, the parts are agents and groups of agents, and the interactions include communication and other kinds of influence. In all these cases, theories are used to make claims about the kinds of mechanisms that produce various kinds of intelligent behavior.

The mechanisms in all these aspects of human mental activity are extraordinarily complex, so cognitive science needs to study them by constructing models, just as occurs in all fields of modern science such as physics. Mathematics is an invaluable tool for describing the interactions between parts, providing general and simplified descriptions of the changes that can result from the interactions. Models differ from theories in providing idealized descriptions of how the mental world is supposed to work. Such idealizations and simplifications make possible detailed explanations and predictions of the results of experiments.

However, to draw out the consequences of a model, it is often crucial to produce a simulation by writing a computer program that implements the mathematical assumptions of the model and the main mechanistic claims of the theory. Generation and testing of a computer program written in LISP, C++, Matlab, or some other higher-level programming language is invaluable in determining whether the idealized parts and interactions actually behave in ways that produce the results expected from observations. I do not mean to suggest that cognitive scientists always proceed from theory to model to program, because the process of discovery often leads in the opposite direction. Thinking about how a program might work can suggest a model that grows into a full-fledged mechanistic theory. But theory, model, and program remain distinct.

I do not expect all cognitive scientists to accept this account of the relation of programs, models, and theories, although I think it gives a good account of decades of practice by myself and others in developing psychological, neural, and organizational theories and programs. More widespread is David Marr's (1982) distinction among computational, algorithm, and implementation levels, which I think is misleading in various respects (e.g., I think that the theories are about mechanisms, not computation). But I will not try to settle the issue here. My crucial point is that the self-understanding of cognitive science cannot simply presuppose any given account of the relations among theories, models, programs, and experiments. These relations have been extensively discussed by philosophers of science, with insights about common practices in physics, biology, and other sciences that need to be applied (or, if necessary) withheld from cognitive science. Ignoring such issues goes hand in hand with simply adopting a philosophical view that may be deeply flawed.

2.3. What are the relations among cognitive science disciplines?

Cognitive science involves at least six integral disciplines: psychology, neuroscience, linguistics, philosophy, anthropology, and artificial intelligence. A major question of the highest generality concerns the relations among and between these disciplines. When cognitive science was officially organized in the late 1970s, the diagram in Fig. 1 served to illustrate actual and possible connections among the six disciplines. An important philosophical problem that should interest all participants in cognitive science concerns the nature of such connections.

One very useful way of thinking about the relations among these disciplines is to think of them as operating at different levels (Churchland & Sejnowski, 1992; Craver, 2007; Newell, 1990). Anthropology operates at the social level, dealing with the interactions among

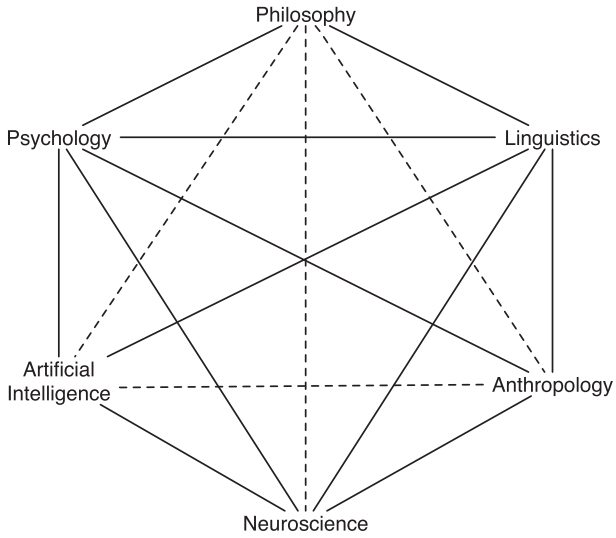

Fig. 1. Connections among the cognitive sciences, based on Gardner, 1985, p. 37, from a 1978 Sloan Foundation report. Unbroken lines indicate strong interdisciplinary ties, and broken lines indicate weak ones. The ties between philosophy and both neuroscience and artificial intelligence are much stronger today.

individuals as part of a culture. Psychology operates at the individual level, concerned with the mental representations and processes of individual thinkers, drawing on ideas from linguistics and artificial intelligence. Psychology, linguistics, and AI can also consider interactions between individuals, as in social psychology, sociolinguistics, and multiagent systems. Neuroscience operates below the psychological level, concerning itself with neural networks. Understanding of neurons often draws also on molecular processes, for example, how genes produce proteins within cells enabling the operations of neurotransmitters such as dopamine and serotonin. What are the relations among operations at these four levels?

Fig. 2 displays four commonly advocated views of such relations. The most familiar is (A), the classical reductionist view that changes at lower levels cause changes at higher levels. It could obviously be taken down to still lower levels involving atoms, subatomic processes, and quantum mechanical effects; but these do not seem relevant to cognitive science as it is currently practiced, so I shall ignore it (see Litt, Eliasmith, Kroon, Weinstein, & Thagard, 2006 for an argument that the brain is not a quantum computer). On this view, causality runs upward and so should explanation: social changes are explained as the result of psychological changes, which are the result of neural changes, all the way down to subatomic changes. This view is far from universally accepted and indeed there are intellectual circles where ''reductionist'' is an epithet almost as vitriolic as ''idiot'' or ''bigot.'' Bickle (2003) unabashedly defends a position he calls ''ruthless reductionism.''

In the social sciences, some writers go far in the other direction, suggesting that the social level is the key source of causality. For example, Karl Marx said that the ideas of the ruling class are in every epoch the ruling ideas, suggesting that psychology is determined to a large extent by economics and sociology. There are many dozens of books written by sociologists

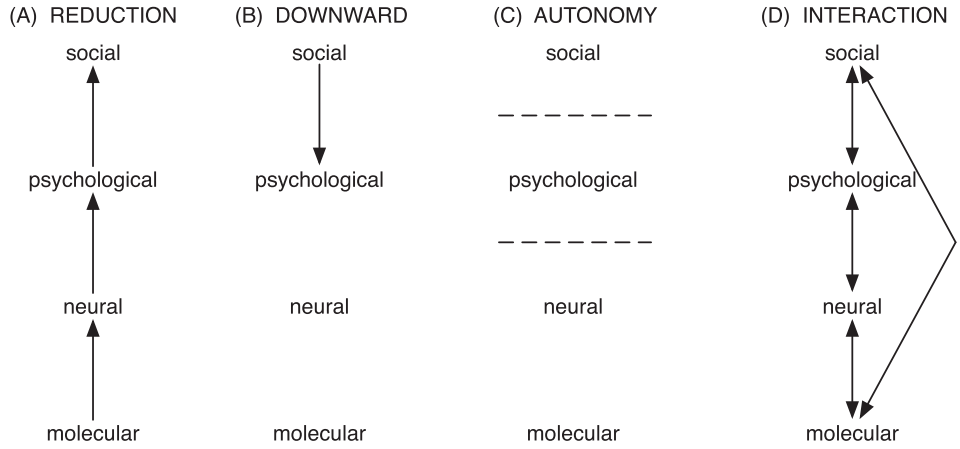

Fig. 2. Four views of the relations between levels of explanation in cognitive science. Arrows indicate causality.

and historians with titles of the form ''the social construction of X'' (Hacking, 1999). Notoriously, Latour and Woolgar (1986) argued that science is a social construction and argued for a moratorium on cognitive explanations of science. On this view, causality and explanation run only downward, from the social to the psychological; the neural and molecular levels are largely ignored.

A more moderate, less imperialistic form of anti-reductionism is the autonomy view, (C) in Fig. 2, where the dotted lines indicate that explanations at each level can proceed independently. This view is popular among sociologists, economists, and anthropologists who want to maintain their independence from psychology without making strong claims of social constructivism. Similarly, some psychologists and philosophers of mind have wanted to defend psychology from the rapidly increasing incursion of neuroscience. One standard argument for defending the autonomy of psychology from neuroscience is the argument from multiple realisability. On this argument, mental states and processes can be instantiated in many different kinds of functional architectures such as robotic ones, not just neurons, so there is an autonomous level of psychological explanation. There have been various philosophical responses to this argument (e.g., Bechtel, 2008; Thagard, 1986) but I think that the most powerful response is just observation of current research trends. Not only cognitive psychology but also social, clinical, and developmental psychology are being increasingly tied to neural processes. Similarly, at the social level, economics is becoming increasingly influenced by behavioral and neural approaches. Hence, the autonomy view is becoming increasingly obsolete.

My own preferred view is the highly interactive one (D), in which there are causal interactions and hence explanatory relations among all levels. This view is not reductionist, because it rejects the one-way causal connections shown in (A), nor is it anti-reductionist, because it recognizes that molecular processes are part of the explanation of neural events, neural processes are part of the explanation of psychological events, and psychological processes are part of the explanation of social events. Elsewhere I defend this multilevel view

in detail with respect to explanation of human emotions (Thagard, 2006) and consciousness (Thagard, forthcoming; Thagard & Aubie, 2008).

Many philosophers and scientists are suspicious of the idea of downward causation as somehow spooky or mystical, but it seems to me unproblematic. Here are just a few examples of cases of the most extreme kind of downward causation, where I think it is legitimate to say that social interactions cause molecular changes:

1. Having to give a presentation increases levels of the stress hormone cortisol.
2. Seeing a beloved causes increased activity of dopamine neurons.
3. Men whose favorite sports team has won a game enjoy increased levels of testosterone.
4. Male chimpanzees who become dominated have lowered levels of testosterone.
5. Women who room together tend to have their menstrual cycles coordinated, altering patterns of estrogen levels.

In short, social changes cause molecular changes.

Craver and Bechtel (2007) argue against the idea of downward causation between levels, but I will not try to respond to their arguments here. The point of my discussion is not to settle the issue, but to show that there is an important general question about levels of explanation in cognitive science that requires philosophical investigation. Ignoring these questions amounts to quiescently adopting one of the four views, usually (A) or (C), without reflection or justification. Progress in cognitive science across the full scope of its ambitions requires assessment of what view of the relations between levels of explanation contributes most to innovative and successful theories and experiments.

2.4. Summary

I have presented only four of the philosophical questions that are crucial to the successful operation of cognitive science, concerning the nature of theories, explanations, computer models, and relations among contributing disciplines. My point has not been simply to assert my own preferred answers to these questions, but to show that there are important controversies that must be addressed as part of a complete, high-level understanding of what cognitive science can accomplish. Ignoring such issues usually amounts to repeating philosophical positions about the nature of scientific knowledge that have proved inadequate in the past, in both the natural and social sciences. Behaviorism, for example, flourished in part because it meshed with the philosophy of positivism, which unduly restricted science to what is observable. The official styles of writing in journals in experimental psychology encode a hypothetic-deductive picture of science that does not fit either with actual practice in psychological research or with currently available philosophical discussions of the relations between science and evidence. In such cases, ignoring philosophy just leads to assumption of persistent but inadequate philosophical views about the general nature of investigation. The best science is highly philosophical because it pays attention to general issues as well as normative ones.

3. Normativity

Science is commonly assumed to be descriptive rather than normative (prescriptive), concerned with how things are rather than with how they ought to be. But applied sciences are also normative in that they aim to make aspects of human life better. For example, clinical psychologists aim to improve the mental health of their patients, and educational psychologists aim to improve learning and teaching. Artificial intelligence researchers aim to build robots that function optimally in their environments. But how can cognitive scientists decide what norms ought to be adopted as the appropriate aims of their research?

Philosophy is the field of cognitive science with most experience in the assessment of normativity. Epistemology, the philosophical theory of knowledge, traditionally addresses whether and how people ought to obtain knowledge. Ethics traditionally addresses how people ought to act. Cognitive science often assumes epistemological and ethical norms without adequate philosophical discussion, as the following issues illustrate.

3.1. Should thinking be Bayesian?

Bayesian models of inference are increasingly popular in cognitive science, with applications as diverse as memory, perception, and neural operations (Anderson, 1990; Griffiths, Kemp, & Tenenbaum, 2008). It is implicitly assumed that, since Bayes' theorem follows from the accepted axioms of probability theory, people ought to be Bayesian, both in their unconscious inferences in perception and memory and also in their conscious deliberations such as theory evaluation, discussed above.

I will not attempt a serious critique of Bayesianism here but merely want to point out that there are philosophical problems that need to be considered before blindly adopting the Bayesian view as most normatively appropriate for cognitive science. The first problem concerns the interpretation of probability: what does it mean to say that the probability of an event (or proposition or whatever) is, say, .4? There are currently three main interpretations of probability, as degrees of belief, as frequencies of occurrence, and as propensities of chance set-ups (Hacking, 2001). The Bayesian view in statistical inference usually assumes the degree of belief interpretation, but Bayesian views of perception seem often to assume the frequency interpretation. I think there are good philosophical grounds for preferring the propensity interpretation. Settling this issue is crucial for assessing whether Bayesian inference really is the appropriate norm to apply to all kinds of inference. I will not attempt to settle it here but merely point out that Bayesianism with respect to various interpretations of probability cannot just be assumed.

Second, even if Bayesian is the appropriate form of inference for data-rich kinds of thinking such as perception and memory, it does not follow automatically that it is the best way to think of high-level conscious inferences about nonobservable states such as mental representations or theoretical entities in science. There are substantial technical and epistemological problems about insisting that scientists and thinkers in general should be Bayesian (see, e.g., Thagard, 2000, ch. 8). Maybe Bayesianism should indeed be the normative view of all

cognitive scientists, but much more discussion is needed before anyone blithely assumes that humans both are and ought to be Bayesian.

Third, Bayesian psychologists assume that cognition is an approximately optimal response to the uncertainty and structure present in natural tasks and environments. Optimality is sometimes justified as arising as the result of natural selection, but the widespread view that natural selection generates optimal solutions has been challenged by biologists and philosophers of biology (e.g., Cowperthwaite, Economo, Harcombe, Miller, & Meyers, 2008; on psychology see Fu, 2008). Hence, the normative question concerning whether people ought to be Bayesian has very interesting psychological, biological, and philosophical dimensions deserving of ongoing discussion.

3.2. How should people make decisions?

Similar issues arise concerning norms for decision making. In the nineteenth century, mathematicians, economists, and philosophers developed the idea that decision makers should maximize expected utility. This view is still dominant in economics, and it often operates in the background of the extensive work in behavioral economics and neuroeconomics that finds people deviating from economic norms. Economists usually supplement utility maximization with game theory in order to consider strategic decisions. There are fundamental issues, however, that need to be addressed in deciding what the norms of decision making should be.

Just as the interpretation of probability is problematic, so is the interpretation of utility (Frey & Stutzer, 2002). In its original form in Bentham and other early theorists, utility was a psychological notion, akin to happiness. Twentieth-century economics moved away from this psychological understanding of utility to a more behaviorist one that interpreted it as a mathematical construction from revealed preferences. More recent work in psychology replaces the behaviorist interpretation with one more akin to the earlier notion, usually under the headings of subjective well-being or just happiness. Attempts have been made to understand estimations of value as a neural process involving the brain's dopamine system (Montague, 2006). Another possibility is that there is no common currency of value that can be summed up as utility or happiness, but rather that there are multiple values that decision making should seek to satisfy (Thagard, forthcoming). I will not attempt to resolve this issue here, but I draw attention to the need for philosophical (and empirical) investigation of the nature of utility that must attend any assumption that the correct norm for decision making involves maximizing expected utility.

It could also be argued that decision making should not aim to maximize or optimize one or more factors, but rather to achieve satisfactory levels as in the satisficing theory of Herbert Simon (1957). Perhaps decision making should be construed, not as a kind of a calculation, but as a sort of parallel constraint satisfaction of the sort performed by neural networks, perhaps involving multiple brain areas (Thagard, 2006). Descriptive studies of decision making such as those involving the ultimatum game, in which people often make decisions based on fairness rather than on maximizing self-interest, also raise questions about what the norms of decision making should be (Hardy-Vallée & Thagard, 2008).

It would take much more argument than I can provide now to show that the standard view of decision making as maximizing expected utility is wrong. But concerns about the nature of utility and the way in which actions ought to be selected as the best way to accomplish goals should slow down any simple view that cognitive science already knows what decision making ought to be.

3.3. Procedures for establishing norms

Because cognitive science cannot dodge the philosophical task of considering how people ought to think, it needs to proceed in a philosophically informed way. One cannot simply assume that some apparently relevant kind of mathematics provides the norms, as I argued concerning the theories of probability and expected utility. Formal logic is another field that might naively be assumed to provide standards for inference, but there is much debate among philosophical logicians concerning which logical standards are appropriate. Some philosophers have thought that pure intuition can provide an a priori basis for establishing inferential standards, but ongoing disagreements among logicians show that a more complex method of establishing norms is needed.

A method of reaching normative conclusions currently popular among philosophers is a method that Rawls (1971) called *reflective equilibrium,* in which people engage in an ongoing process of adjusting their inferential norms based on their practices and their practices based on their norms. This method has numerous problems, however, including undue reliance on intuition and the danger of arriving at stable but suboptimal sets of norms (Harman & Kulkarni, 2007; Thagard, 1988, 2000). Here is a more consequentialist method of establishing norms defended by Thagard (forthcoming):

1. Identify a domain of practices.
2. Identify candidate norms for these practices.
3. Identify the appropriate goals of the practices in the given domain.
4. Evaluate the extent to which different practices accomplish the relevant goals.
5. Adopt as domain norms those practices that best accomplish the relevant goals.

The tricky part of this method is step 3, identifying the appropriate goals, which seems to be as much a normative as a descriptive matter. However, it can at least have a large empirical component, as when we study people's decision-making behavior to determine what goals they pursue and use psychology and neuroscience to understand why they pursue those goals. For example, I think that love, work, and play are more appropriate goals of decision making than money, because they help people satisfy vital needs, which are much more biologically basic than wants and preferences (Thagard, forthcoming).

3.4. Summary

I have not said nearly enough to justify my favored schema for establishing norms, let alone to justify my preferred norms for inductive inference, decision making, and ethical

behavior. But I hope to have indicated that cognitive science cannot take for granted common norms such as Bayesian inference, expected utility theory, and game theory. Deciding what norms are appropriate, and deciding how to decide what norms are appropriate, are philosophical problems that the more empirical fields of cognitive science cannot tackle on their own. In normative matters, ignoring philosophy amounts to doing it implicitly, and badly.

4. Why philosophy needs cognitive science

Hence, cognitive science needs philosophy, but it does not need most styles of philosophy. Here are some approaches that are inimical to the aims and methods of cognitive science:

1. Rationalist approaches that assume philosophical truths can be gained by reason alone (e.g., Plato, Kant, Hegel).
2. Analytic approaches that rigidly divorce thought from thinking and assume that formal logic and/or linguistic analysis are the best tools for investigating knowledge (e.g., Frege, Dummett).
3. Postmodernist approaches that disparage science and toss scientific and philosophical terms around with careless abandon (e.g., Deleuze, Derrida).

Cognitive science can gain scant benefit from approaches that view philosophy as independent from and superior to scientific investigation, although there may be occasional theoretical ideas such as Kant's theory of schemas and Frege's theory of relations that prove scientifically useful.

Fortunately, there is a venerable history of naturalist approaches to philosophy that ties it closely to science, with such distinguished practitioners as Aristotle, John Locke, David Hume, John Stuart Mill, Charles S. Peirce, William James, John Dewey, and W. V. O. Quine. Examples of recent philosophical works espousing or exemplifying naturalism include Appiah (2008); Bechtel (2008); Bunge (2006); Churchland (2007); Churchland (2002); Craver (2007); Goldman (2006); Harman and Kulkarni (2007); Knobe and Nichols (2008); Maddy (2007); Nersessian (2008); Sinnott-Armstrong (2008); and Thagard (2006, forthcoming). Naturalistic philosophy is far from monolithic and there are many points of disagreement concerning scientific as well as philosophical issues. But naturalists agree that progress in philosophy requires close attention to scientific developments. Whereas rationalist, analytical, and postmodernist approaches resist the incursion of scientific findings into philosophy, naturalistic approaches welcome them as highly relevant to the generality and normativity concerns of philosophy that I discussed in the last section.

From a naturalistic perspective, the common philosophical practice of using thought experiments to clarify concepts and arrive at a priori truths is inherently suspect (Thagard, forthcoming; but for defense of thought experiments, see Shepard, 2008; and Williamson, 2007). The idea that philosophy has a special role in clearing up conceptual confusion has been highly popular in the last century, but experimental work on concepts shows it to be

problematic (see Murphy, 2002 for an excellent survey). As Quine (1963) argued on purely theoretical grounds and as subsequent experimental work has confirmed, concepts are intimately tied to beliefs and theoretical explanations of the world. Concepts change as scientific theories change, not because philosophers concoct fanciful cases to pump intuitions to defend their own preferred versions of concepts. The philosophical idea of conceptual confusion is a confusion about concepts, whose clarification needs to be part of the scientific enterprise of developing better theories, not a stand-alone linguistic exercise.

Thought experiments can be useful tools in developing hypotheses and contrasting them with alternative ones, but they are useless for justifying the acceptance of hypotheses. Without additional scientific evaluation, thought experiments are more likely to yield falsehoods than truths, let alone necessary truths. The history of science shows that concepts change radically over time in response to theoretical and experimental developments, bringing about major reorganizations in views of what kinds of things there are (Thagard, 1992, 2008). Hence, the philosophical category of category mistakes is a mistake about categories.

Even if conceptual investigation could show that some beliefs are innate, that would not be grounds for assuming them to be true. Naturalistically, innate beliefs are just ones that became genetically established in human brains through natural selection, which is not guaranteed to arrive at truths, only at beliefs that can be useful for survival and reproduction. For example, the empirically supported theory of general relativity suggests limitations on what may be an innate human expectation that the world conforms to Euclidean geometry.

The limitations of rationalist, analytic, and postmodernist philosophy should make it clear why all the major areas of philosophy need cognitive science. To establish general hypotheses and normative conclusions about the nature of mind, knowledge, and morality, philosophy needs to take into account a wide range of information from relevant findings in psychology, neuroscience, linguistics, and so on. Such information not only provides philosophers with content to philosophize about but also constraints on what kinds of theories they can reasonably defend. Cognitive science is relevant to philosophy not just in narrow fields such as philosophy of mind and philosophy of science, but at the most general level in metaphysics, epistemology, and ethics (Thagard, forthcoming).

5. Conclusion

I have argued that cognitive science needs philosophy in its pursuit of answers to general and normative questions. Ignoring scientifically informed philosophical reflection leads not only to bad philosophy but also to bad science. Positivist philosophy of science limits cognitive science to behaviorism; idealist and relativist metaphysics and epistemology incline cognitive science to postmodernism and social constructivism. Neglect of reflection about ethical principles generates the default undergraduate view that morality is subjectively relative to cultures and individuals.

The proper role of philosophy in cognitive science can be illuminated by considering various metaphors that philosophers have used to characterize their enterprise (Thagard &

Beam, 2004). Many philosophers such as Descartes have thought that philosophy ought to provide foundations for knowledge, which would require it to be prior to science:

> Throughout my writings I have made it clear that my method imitates that of the architect. When an architect wants to build a house which is stable on ground where there is a sandy topsoil over underlying rock, or clay, or some other firm base, he begins by digging out a set of trenches from which he removes the sand, and anything resting on or mixed in with the sand, so that he can lay his foundations on firm soil. In the same way, I began by taking everything that was doubtful and throwing it out, like sand; and then, when I noticed that it is impossible to doubt that a doubting or thinking substance exists, I took this as the bedrock on which I could lay the foundations of my philosophy. (Descartes, 1984, vol. 2, p. 366)

If science needed foundations, then it would be a legitimate job for philosophy to provide them.

In contrast, naturalistic approaches are more consistent with a view of knowledge, not as a building with foundations, but rather as a cable (Peirce, 1958, pp. 40–41):

> Philosophy ought to imitate the successful sciences in its methods, so far as to proceed only from tangible premisses which can be subjected to careful scrutiny, and to trust rather to the multitude and variety of its arguments than to the conclusiveness of any one. Its reasoning should not form a chain which is no stronger than its weakest link, but a cable whose fibers may be ever so slender, provided they are sufficiently numerous and intimately connected.

Another more recent coherentist metaphor is Neurath's (1959, p. 201) ship:

> There is no way of taking conclusively established pure protocol sentences as the starting point of the sciences. No *tabula rasa* exists. We are like sailors who must rebuild their ship on the open sea, never able to dismantle it in dry-dock and to reconstruct it there out of the best materials. Only the metaphysical elements can be allowed to vanish without trace. Vague linguistic conglomerations always remain in one way or another as components of the ship.

Unlike Descartes' foundations, Peirce's cable and Neurath's ship not only allow but require the enterprise of philosophy to be entwined with all of science, including cognitive science.

My all-time favorite analogy for philosophy and for science is from Francis Bacon, writing around 1620 when the two enterprises had not yet been distinguished (Bacon, 1960, p. 93):

> Those who have handled sciences have been either men of experiment or men of dogmas. The men of experiment are like the ant; they only collect and use. The reasoners resemble

> spiders who make cobwebs out of their own substance. But the bee takes a middle course. It gathers its material from the flowers of the garden and of the field, but transforms and digests it by a power of its own. Not unlike this is the true business of philosophy; for it neither relies solely or chiefly on the powers of the mind, nor does it take the matter which it gathers from natural history and mechanical experiments, and lay it up in the memory whole as it finds it, but lays it up in the understanding altered and digested. Therefore, from a closer and purer league between these two faculties, the experimental and the rational (such as has never yet been made), much may be hoped.

Similarly, from a closer and purer league between cognitive science and philosophy much may be hoped.

Acknowledgments

Thanks to Michael Anderson, Andrew Brook, and Wayne Gray for helpful comments on an earlier draft. Research support was provided by the Natural Sciences and Engineering Research Council of Canada.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson, J. R. (2007). *How can the mind occur in the physical universe?* Oxford, England: Oxford University Press.

Appiah, A. (2008). *Experiments in ethics*. Cambridge, MA: Harvard University Press.

Bacon, F. (1960). *The New Organon and related writings*. Indianapolis, IN: Bobbs-Merrill.

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. New York: Routledge.

Bechtel, W., & Richardson, R. C. (1993). *Discovering complexity*. Princeton, NJ: Princeton University Press.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer.

Bunge, M. (2006). *Chasing reality: Strife over realism*. Toronto: University of Toronto Press.

Churchland, P. S. (2002). *Brain-wise: Studies in neurophilosophy*. Cambridge, MA: MIT Press.

Churchland, P. (2007). *Neurophilosophy at work*. Cambridge, England: Cambridge University Press.

Churchland, P. S., & Sejnowski, T. (1992). *The computational brain*. Cambridge, MA: MIT Press.

Cowperthwaite, M. C., Economo, E. P., Harcombe, W. R., Miller, E. L., & Meyers, L. A. (2008). The ascent of the abundant: How mutational networks constrain evolution. *PLoS Computational Biology*, *4*(7), e1000110.

Craver, C. F. (2007). *Explaining the brain*. Oxford, England: Oxford University Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *22*, 547–663.

Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge, England: Cambridge University Press.

Descartes, R. (1984). *Philosophical writings of descartes*, J. Cottingham, R. Stoothof & D. Murdoch (Trans). Cambridge, England: Cambridge University Press.

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation and dynamics in neurobiological systems*. Cambridge, MA: MIT Press.

Frey, B. S., & Stutzer, A. (2002). *Happiness and economics*. Princeton, NJ: Princeton University Press.

Fu, W. (2008). Is a single-bladed knife enough to dissect human cognition? Commentary on Griffiths et al. *Cognitive Science*, *32*, 155–161.
Gardner, H. (1985). *The mind's new science*. New York: Basic Books.
Goldman, A. I. (2006). *Simulating minds*. New York: Oxford University Press.
Griffiths, T. L., Kemp, C., & Tenenbaum, J. B. (2008). Bayesian models of cognition. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 59–100). Cambridge, England: Cambridge University Press.
Hacking, I. (1999). *The social construction of what?* Cambridge, MA: Harvard University Press.
Hacking, I. (2001). *An introduction to probabiliity and inductive logic*. Cambridge, England: Cambridge University Press.
Hardy-Vallée, B., & Thagard, P. (2008). How to play the ultimatum game: An engineering approach to metanormativity. *Philosophical Psychology*, *21*, 173–192.
Harman, G., & Kulkarni, S. (2007). *Reliable reasoning: Induction and statistical learning theory*. Cambridge, MA: MIT Press.
Hempel, C. G. (1965). *Aspects of scientific explanation*. New York: The Free Press.
Knobe, J., & Nichols, S. (2008). *Experimental philosophy*. Oxford, England: Oxford University Press.
Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
Litt, A., Eliasmith, C., Kroon, F. W., Weinstein, S., & Thagard, P. (2006). Is the brain a quantum computer? *Cognitive Science*, *30*, 593–603.
Maddy, P. (2007). *Second philosophy: A naturalistic method*. Oxford, England: Oxford University Press.
Marr, D. (1982). *Vision*. San Francisco: Freeman.
Montague, R. (2006). *Why choose this book? How we make decisions*. New York: Penguin.
Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
Nersessian, N. (2008). *Creating scientific concepts*. Cambridge, MA: MIT Press.
Neurath, O. (1959). Protocol sentences. In: A. J. Ayer (Ed.), *Logical positivism* (pp. 199–208). Glencoe, IL: The Free Press.
Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
Newell, A., Shaw, J. C., & Simon, H. (1958). Elements of a theory of human problem solving. *Psychological Review*, *65*, 151–166.
Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
Peirce, C. S. (1958). *Charles S. Peirce: Selected writings*. New York: Dover.
Popper, K. (1959). *The logic of scientific discovery*. London: Hutchinson.
Quine, W. V. O. (1963). *From a logical point of view* (2nd ed.). New York: Harper Torchbooks.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Rumelhart, D. E., & McClelland, J. L. (Eds.). (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge MA: MIT Press/Bradford Books.
Shepard, R. N. (2008). The step to rationality: The efficacy of thought experiments in science, ethics, and free will. *Cognitive Science*, *32*, 3–35.
Simon, H. (1957). *Models of man: Social and rational*. New York: Wiley.
Sinnott-Armstrong, W. (Ed.). (2008). *Moral psychology (3 volumes)*. Cambridge, MA: MIT Press.
Sober, E. (2008). *Evidence and evolution: The logic behind the science*. Cambridge, England: Cambridge University Press.
Thagard, P. (1986). Parallel computation and the mind-body problem. *Cognitive Science*, *10*, 301–318.
Thagard, P. (1988). *Computational philosophy of science*. Cambridge, MA: MIT Press/Bradford Books.
Thagard, P. (1992). *Conceptual revolutions*. Princeton, NJ: Princeton University Press.
Thagard, P. (1999). *How scientists explain disease*. Princeton, NJ: Princeton University Press.
Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.
Thagard, P. (2005). *Mind: Introduction to cognitive science* (2nd ed.). Cambridge, MA: MIT Press.

Thagard, P. (2006). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge, MA: MIT Press.

Thagard, P. (2008). Conceptual change in the history of science: Life, mind, and disease. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 374–387). London: Routledge.

Thagard, P. (forthcoming). *Brains and the meaning of life*. Princeton, NJ: Princeton University Press.

Thagard, P., & Aubie, B. (2008). Emotional consciousness: A neural model of how cognitive appraisal and somatic perception interact to produce qualitative experience. *Consciousness and Cognition*, *17*, 811–834.

Thagard, P., & Beam, C. (2004). Epistemological metaphors and the nature of philosophy. *Metaphilosophy*, *35*, 504–516.

Thagard, P., & Litt, A. (2008). Models of scientific explanation. In R. Sun (Ed.), *The Cambridge handbook of computational psychology* (pp. 549–564). Cambridge, England: Cambridge University Press.

Williamson, T. (2007). *The philosophy of philosophy*. Malden, MA: Blackwell.

Wittgenstein, L. (1971). *Tractatus logico-philosophicus*, (2nd ed.) D. F. Pears & B. F. McGuinness (Trans). London: Routledge & Kegan Paul.