Epistemology, Rationality, and Cognition

John L. Pollock
Department of Philosophy
University of Arizona
Tucson, Arizona 85721
pollock@arizona.edu
http://www.u.arizona.edu/~pollock

1. An Architecture for Rational Cognition

1.1 Setting Aside the Gettier Problem

Since Gettier, much of epistemology has focused on analyzing "S knows that P", but that is not my interest. My general interest is in rational cognition — both in what it is to be rational, and in how rational cognition works. The traditional epistemological question, "How do you know?", can be taken as addressing part of the more general problem of producing a theory of rational cognition. It is about specifically epistemic rationality. I interpret this question literally, as a question about *how* we should proceed in our epistemic endeavors. Epistemological theories that try to answer this question are theories of *procedural epistemology* (see my 1998), and when, from this perspective, we assess beliefs in terms of their justifiedness, the concept of justification is one of *procedural epistemic justification*. Whether this has anything to do with the analysis of knowledge is an open question, and not one that I have much interest in addressing.

1.2 Interest Driven Epistemic Cognition

I think it is helpful to approach epistemology from the design stance, and ask what role rational epistemic cognition has in the broader cognitive architecture of a cognitive agent. We can make a rough division of rational cognition into epistemic cognition, which is about what to believe, and practical cognition, which is about what to do. Epistemologists normally assume that we can study epistemic cognition without paying any attention to practical cognition. But approaching epistemology from the broader perspective of designing a rational agent quickly gives the lie to this assumption. The main point of a system of cognition is to direct an agent's interaction with the world. So its main purpose is to direct action. The intelligent direction of action requires the agent to have information both about its environment and about itself. It is the function of epistemic cognition to provide that information. Viewed in this way, epistemic cognition is subservient to practical cognition. Its role is to provide the information needed for rational decision making. This leads immediately to one important characteristic of epistemic cognition. When we reason, we do not reason at random, drawing conclusions willy nilly as they come to us. Rather, we seek to answer specific questions, derived initially from queries posed by practical cognition.[1] For example, if I learn that my favorite author has written a new book, I may adopt the goal of reading it.

[1] See my (1995) for a fuller discussion of *interest-driven reasoning*.

Practical cognition then poses the problem of how to achieve this goal, and passes a query to epistemic cognition, asking for potential plans for achieving it. Epistemic cognition engages in reasoning aimed at the production of plans for achieving the goal, and if plans are found, they are passed back to practical cognition, which then adopts interest in evaluating them it. It may be able to do that by appealing to values stored in an evaluative database (see my 2006 for details), or it may have to acquire more information in order to assess the expected utilities of the plans. In the latter case, queries aimed at acquiring that information are passed back to epistemic cognition. So epistemic cognition and practical cognition talk to each other, and much of our epistemic cognition is driven by prior practical cognition. I will put this by saying that epistemic cognition is "interest driven". I gave an account of how this works in my (1995).

It might seem that perception constitutes an exception to the rule that epistemic cognition is driven by practical cognition. Philosophers are sometimes attracted by the simplistic view that in perception we just take in information as it is presented to our senses. However, this is readily shown to be false. Consider vision. Visual perception produces a very rich visual image, and it requires attention to retrieve information from it for further processing. A now familiar illustration of this is Simons and Chabris' example (1999) of the gorilla in the basketball game. You have to see this example to fully appreciate it, and if you want to experience it from a first-person point of view, go to http://viscog.beckman.uiuc.edu/media/ig.html to see it for yourself. Check it out *before* reading on, or the example will be spoiled for you. In this example, subjects see a video clip of a group of eight students passing a basketball around, and the subjects are asked to keep track of how many times the ball is passed from one student to another. The game goes on for a couple of minutes. Midway through the game, another student dressed in a gorilla suit walks through the middle of the game. Afterwards, the subjects are asked whether they saw the gorilla. Most of them say, "What gorilla?" When they are shown the video again, they look for and see the gorilla. Some of them have to be convinced that it is the same video, because they cannot believe they would have overlooked it.

For a simpler example, scan a group of people (e.g., the students in a small class). Afterwards, if you are asked what color shirt Jim was wearing, you may not know. You did not "notice it", although a perceptual representation of it was present in your visual image. The point of both of these examples is that belief formation based on the perceptual presentation of information is profoundly influenced by your interests. You must attend to the information before you can form beliefs about it. Some mechanism for attention are automatic. For example, sudden motions in an otherwise still scene, or flashing lights, grab your attention automatically. But many mechanisms of attention are interest driven. If I asked you what color Jim's shirt is while you are still looking at the class, you will be able to attend to it and answer the question.

1.3 Empirical Investigation

That epistemic cognition is interest driven is the simplest way in which it is influenced by practical cognition. Another fairly obvious connection that has nonetheless been ignored in most

epistemological investigations is that epistemic queries cannot usually be answered simply by reasoning from information the agent already has. In the preceding example, when asked what color Jim's shirt is, you have to look — you cannot just close your eyes and reason about it a priori. The acquisition of new information sought for practical purposes typically involves some degree of "empirical investigation", ranging from simply directing your attention, to redirecting your eyes, to looking up information up in a book or online, to engaging in scientific experiments. These are all actions that you perform, and as such are driven by practical cognition. You often have to engage in difficult problem solving (practical cognition) in deciding how to pursue a desired piece of information. So epistemic cognition often initiates new practical cognition, driven by "epistemic desires" for specific information. This leads practical cognition to pose a question to epistemic cognition regarding how to acquire the information. That may lead to further epistemic cognition, which may initiate further practical cognition, and so on. In other words, rational cognition incorporates loops between epistemic and practical cognition. Neither can accomplish much without the other.

1.4 Reflexive Cognition

These interactions between practical cognition and epistemic cognition are all fairly obvious, and the traditional epistemologist may retort that of course there are interactions, but his interest is in the purely epistemic cognition that transpires once practical cognition has posed its queries. However, this is still a simplistic view of epistemic cognition. The difficulty is that human beings are *reflexive cognizers*. We do not just engage in cognition about the world. We also engage in cognition about cognition, and that often gives us the power to redirect the course of our own cognition.[2] For example, I may be faced with two problems: (1) finding a unified field theory for physics; (2) where to go to lunch. Typically, we are faced with more cognitive tasks than we can immediately undertake, so we need some way of prioritizing them. We can regard these prioritized tasks are stored on a "cognitive task queue", and retrieved in order of priority. There has to be a way of prioritizing them automatically, without thinking about it, because otherwise we would be led into an infinite regress. Plausibly, our default ordering of problem-solving tasks would order the more important ones before the less important ones. Finding a unified field theory is more important than deciding where to go the lunch, so that would take priority. However, a sensible cognizer will also recognize that although that is the more important problem, it is also one he is less likely to solve in the immediate future, and so he may decide to put it aside for a bit and go to lunch. This involves reasoning explicitly about what problems to address and in what order. This is something over which we, as reflexive cognizers, have control. We can decide what to think about. Similarly, in trying to solve a purely epistemic problem, a cognizer may consider alternative strategies and decide which to pursue first, again explicitly altering the ordering of his cognitive task queue. The ability to do this is very important for efficient problem-solving, even when the problems are purely epistemic. But now notice that the reasoning involved here is practical

[2] See my (2007) for a more extensive discussion of this. See also my (1995) and Pollock and Cruz (2000).

cognition. Re-ordering my cognitive task queue is *something I do*, and the reasoning involved is reasoning about whether or how to do it. As such it appeals to the same kind of decision-theoretic considerations as any other practical reasoning.

There are a number of other ways in which we can alter the course of our own cognition. For example, we can decide to refrain from accepting the conclusion of some bit of reasoning even when we are firmly convinced that the premises are true and we cannot see anything wrong with the argument. An obvious example of this occurs when we are presented with logical paradoxes. For instance, the liar paradox leads us to a contradiction. Few people think they know what is wrong with the reasoning that leads to the contradiction, but they do not accept the conclusion that the contradiction is true. Instead, they are able to back out of the problem and quit thinking about it. They may return to it later, perhaps repeatedly, but without ever solving it they can still refrain from accepting the conclusion. This same phenomenon occurs elsewhere. Presented with an argument for the non-existence of God, someone may refuse to accept the conclusion even though he cannot say what is wrong with the argument. Someone else, presented with an argument for the existence of God, may equally refuse to accept that conclusion, although he cannot say what is wrong with that argument either. The ability to refrain from accepting an argument is an important one. We often have to think long and hard about an argument to figure out what is wrong with it. When we suspect something is wrong with an argument, we do not want to have to accept the conclusion until we later find the error in the argument. We want to be able to set the argument aside, perhaps to return to it later, perhaps to ignore it forever.

Another aspect of reflexive cognition that will turn out below to have important implications for philosophical methodology is that we often accept conclusions on the basis of incomplete argument sketches rather than fully worked out arguments. Consider mathematical theorem proving. Philosophers often have the fantasy that that produces the most certain of knowledge. In fact, mathematical reasoning is one of the most error prone forms of reasoning, and professional mathematicians adopt a healthy skepticism towards new arguments purporting to establish novel conclusions. When first produced, complex mathematical proofs almost always contain errors. Sometimes the errors can be repaired, and sometimes they cannot. Part of the explanation for this is that the "proofs" are really argument sketches rather than complete arguments. No one produces fully worked out arguments of the sort we teach students to produce in introductory logic courses. Such arguments would be too long to comprehend — we would lose track of the forest for the trees. Instead, mathematicians produce argument sketches, with the presumption that the details could all be filled in as necessary. Their basis for believing this is probably some sort of reasoning by analogy from previous cases. This is reflexive cognition, because we are reasoning about our reasoning and concluding that we could fill in the details if we had to.

It should be emphasized that the use of argument sketches is not the exclusive province of mathematics. Everyone forms beliefs on the basis of argument sketches. E.g., one person might reason that because there is so much evil in the world, there cannot be a benevolent God. Someone else might reason that there must be a God because the world exhibits intelligent design. These are

argument sketches, and they are notoriously difficult to fill out.

The upshot of the preceding observations is that epistemic cognition and practical cognition are not separable modules. Only the lowest level of epistemic cognition can proceed without the intervention of practical cognition, and even then our epistemic pursuits are interest-driven. A complete theory of epistemic rationality is inextricably interwoven with a theory of practical rationality. One implication of this is that familiar attempts to characterize epistemic cognition as a goal-directed activity, the goal being the acquisition of true beliefs, is wrong-headed. An architecture for epistemic cognition cannot be evaluated independently of its interactions with practical cognition. They jointly form a cognitive architecture, and what makes the epistemic parts of it good or bad is how they contribute to the functioning of the whole architecture. This cannot be evaluated by anything so simple as its propensity to produce true beliefs. At the very least, epistemic cognition must produce beliefs that are useful to agent, and what makes them useful is their role in facilitating the solution to practical problems. It is not obvious that beliefs must be literally true for this purpose, and it is clear that merely being true is not enough to make beliefs useful.

1.5 A Two-Factor Architecture

A very important feature of the human cognitive architecture, and probably an essential feature of any cognitive architecture able to function efficiently in a complex and rapidly changing environment, is that beliefs and decisions need not be the product of explicit reasoning. Suppose I toss an apple to you and you catch it. How did you do that? You certainly did not do it by measuring distances and velocities and computing parabolic trajectories. Perhaps you *could* have done it that way, but it would have been much too slow and you would not have caught the apple. Instead, humans and most higher animals have a built-in cognitive module whose purpose is to rapidly produce predictions of trajectories. We rely upon that in forming beliefs about where the apple is going to be when we try to catch it. I call such modules *Q&I modules* ("quick and inflexible modules", from Pollock 1989).[3]

I have long emphasized the importance of Q&I modules in epistemic cognition. For example, it seems that most inductive and probabilistic beliefs are produced by Q&I modules. The general problem is that explicit reasoning is slow and computationally difficult. Q&I modules produce beliefs quickly, but they do so by making assumptions that may fail. For instance, our trajectory module only works insofar as the flight of the object is unimpeded. As such, a cognitive agent must be able to discover when Q&I modules are apt to produce incorrect results and override them in those circumstances. Giving priority to explicit reasoning over Q&I modules is an essential feature of our cognitive architecture.

Q&I modules are important for epistemic cognition, but they may be even more important for understanding human practical cognition. In recent years I have become more and more a

[3] See my (2006) for a more extensive discussion of Q&I modules and their role in both epistemic and practical cognition.

confirmed nativist. I suspect that most human decision making, particularly in social contexts, is carried out by Q&I modules. Inborn personality traits reflect individual differences in these practical Q&I modules. Again, when we have the relevant information to make informed decisions on the basis of reasonably held beliefs about values and probabilities, rationality dictates that we should override our Q&I modules and engage in some form of decision-theoretic reasoning (see my 2006 for a detailed account of this). But social situations are often so complex that it is extremely difficult to reason about them explicitly, in part because we lack knowledge of the relevant probabilities and utilities. And other kinds of practical decision making, e.g., predator avoidance (which also occurs in social situations) may require very rapid decision making, which cannot wait for an exhaustive decision-theoretic analysis.

The result is a two-factor cognitive architecture, in which most beliefs and decisions are produced quickly and fairly automatically by Q&I modules. Explicit reasoning sits above this bundle of Q&I modules and (1) attempts to modulate it, overriding conclusions that can be expected to be wrong, and (2) tries to fill in the gaps when our Q&I modules are unable to produce automatic solutions to problems. For example, my Q&I modules are good at telling me when to stop working and go to lunch, but they will never produce a unified field theory. For the latter we need careful scientific investigation and lots of explicit reasoning (although even there Q&I modules play essential roles in "micro-decision-making").

1.6 Rules of Rationality

Epistemologists often suppose that they are in the business of producing rules for rational cognition. Simple examples of such rules might be "Do not hold a belief for which you have no good reason", "Accept the conclusion of a good argument", and more complicated rules would tell us when to accept the conclusions of inductive arguments, how handle probabilities, when to draw conclusions on the basis of perceptual input, etc. But the preceding conclusions demonstrate that none of these rules can be right. Reflexive cognition can lead us to refrain from accepting the conclusion of an argument even when we cannot see anything wrong with it, or to accept a conclusion on the basis of an argument sketch, without having a complete argument. Such cognitive behavior is not irrational. And the same point applies to the more complex principles describing inductive or probabilistic reasoning and perception. These are at best default rules for how to proceed in the absence of reflexive cognition.[4] True rules of rationality must describe how these default rules interact with reflexive cognition, and stating such rules probably requires formulating a complete theory of rational cognition. It is unlikely that there are piecemeal rules that have any standing on their own, independently of being embedded in the larger system.

2. Human Rationality

[4] See my (2007) for a more careful discussion of this.

Investigating rationality from the design stance can take either of two directions. We can ask how to build a well-functioning cognitive agent, without caring much whether it works the way humans work. This is one strand of research in artificial intelligence. But we can also ask specifically how human rational cognition works, and employ the design stance in an attempt to understand why it works in the way it does. The design stance imposes important constraints on theories of human rationality generally, and on epistemological theories specifically, because they are theories about *how we work*. If such a theory is to be correct, it must be possible for an agent to actually work that way. Thus an important test of such a theory is to consider whether, if we built an agent that worked in the way envisioned, it would perform in ways that we regard as rational. Would it accord with our standards of rationality? If not, the theory must be wrong.

Epistemologists have traditionally tried to meet this requirement by engaging in thought experiments run from their armchairs, and that must inevitably be the first test of an epistemological theory. But we are very limited as to how far we can go in this way. The problem is that, as in all philosophy, the devil is in the details. It is easy to sketch theories that look good in the abstract, but they usually break down because it is impossible to fill in the details in any reasonable way. In epistemology, the only way to surmount this problem is to work out the details. Philosophers are unaccustomed to working this hard, partly because once the details have been supplied, the theory tends to become too complex to evaluate from the armchair. This is for two reasons: (1) It can be difficult to be sure what consequences a complex theory has when applied to a complicated case: (2) It can be surprisingly difficult to be sure that you have even supplied all the details needed for the theory to have any implications at all. Both of these problems are familiar to computer programmers. No matter how skillful the programmer, a complex computer program *never ever* works properly the first time. All complex programs are initially buggy, and the only way to find the bugs is to run the program on test cases and see what it does. Epistemological theories are much like computer programs. They are theories about *how something works*, namely, rational cognition, and the only way to get such a theory right is to run it on complex cases. If the theory is complex and the cases are complicated, we cannot do this by just sitting in our armchairs and thinking. As far as I can see, the only way to do it is to actually build an agent that works in the way described by the theory and see what it does. In other words, we must test theories of rational cognition by building AI systems that model them. This conviction gave rise to the OSCAR project in 1985.[5] The objective of the OSCAR project is to construct of a general theory of rationality and test it by implementing it as an AI system. We can then apply the working system to complex scenarios and see what it does. As with any computer program, I can assure you that no implemented theory of rational cognition will do what we expect it to do the first time around. Assuming that the program faithfully captures the theory, designing and refining the theory must proceed as programming always does — by systematic testing and debugging. In this case it is the theory itself that we are debugging.

[5] For a continually updated report on the status of OSCAR, see http://oscarhome.soc-sci.arizona.edu/ftp/OSCAR-web-page/oscar.html.

In epistemology, however, there is another twist to the problem. In the hard sciences, we test our theories by looking at the world and seeing whether it behaves in the manner portrayed by the theory. In philosophy, the criterion for correctness of a theory of rational cognition is that a system performing as described will actually behave rationally. But to test a theory in this way, we must be able to tell whether a concrete bit of behavior does accord with our standards of rationality. How do we do that? Epistemologists have traditionally tested their theories by appealing to their "philosophical intuitions". But what are these, and why should we trust them?

Although theories of privileged access have fallen into disrepute in the philosophy of mind, I claim that we do have a kind of privileged access to the rationality of our judgments. This reflects an important feature of our cognitive architecture. I remarked above that we often form beliefs on the basis of argument sketches rather than complete arguments. But for this to work satisfactorily, we must be able to criticize argument sketches on the grounds that there is no way to fill them out into complete arguments, or alternatively we must be able to confirm an argument sketch by showing that it can be filled out. To do that, we must be able to inspect candidate expansions of argument sketches and evaluate them as good or bad arguments. But that just amounts to judging whether, if we reasoned in that way, we would be conforming to the dictates of rationality. Thus an essential feature of rational cognition must be the built-in ability to judge whether particular bits of cognitive behavior conform to the dictates of rationality. This kind of self-monitoring of cognition is perfectly analogous to our similar ability to monitor physical behavior and judge whether we are performing our actions in conformance with our procedural knowledge for how to do whatever it is that we are doing.[6] This is a built-in feature of our cognitive architecture, and it is this that underlies our so-called "philosophical intuitions", at least regarding rational cognition, and in particular, regarding epistemic cognition.

The preceding explains how we can judge the rationality of particular bits of cognitive behavior, but what is it *to be rational*? By virtue of what are the built-in standards to which our philosophical intuitions appeal correct? I doubt that this has an answer, at least in the form of a logical analysis of rationality. Judging rationality is a built-in feature of reflexive cognition, and thus just one more aspect of our cognitive architecture. The ability to make these judgments is included in the architecture because it makes the complete architecture work better. In saying this, we are judging the architecture from an external perspective, and there are numerous external perspectives from which we might make such an evaluation. E.g., we might adopt an evolutionary perspective and ask how well the architecture contributes to the propagation of the human genome. This is a different kind of judgment than our internal judgments of rationality. They are part and parcel of the system itself, and proceed as they do because that is the way we are built. To be rational just is to conform to our reflective judgments of rationality. I doubt that there is more to be said on the matter. From an external perspective, we may be able to show why a system whose rational standards satisfy some broad general principles will tend to work better than one that does not. For instance, I think it can be argued convincingly that any sophisticated cognitive agent needs a system

[6] This proposal derives originally from my (1987).

of defeasible reasoning.[7] But there may be many different systems of reasoning that will work equally well from an external perspective. Only one of these will be endorsed by the agent's internal judgments of rationality, but which one that is may be largely an accident. For example, psychologists have argued convincingly that *modus tollens* is not a built-in inference rule for human beings (Wason 1996; Cheng and Holyoak 1985). It is claimed that initially we get the same result by using *contraposition* and *modus ponens*. Later, we may learn the rule of *modus tollens* and employ it in our reasoning as a derived rule. If this is correct, then our built-in standards of rationality would deem reasoning by *modus tollens* irrational until the agent has learned it on some other basis. But surely, an agent with a cognitive architecture that differs from ours only in having *modus tollens* as an additional built-in inference rule would not thereby be an inferior agent from an external perspective.

3. Human Irrationality

Epistemologists sometimes regard their task as that of discovering (or discovering how to discover) rules for avoiding irrationality. But they rarely stop to consider a deeper question. Why is is possible for human beings to be irrational? If evolution has deemed it desirable for us to behave according to certain standards of rationality, why didn't it just build us so that we work that way? For example, consider my artificial agent OSCAR. OSCAR is able to engage in some quite sophisticated cognition, but (unless it is broken) OSCAR cannot be irrational. OSCAR is built to work in accordance with rules motivated by studies of human rationality, but OSCAR cannot violate those rules. That is just the way OSCAR works. Humans, on the other hand, can behave irrationally without being broken. Why is this possible?

What makes human irrationality possible is that humans are reflexive cognizers. Unlike OSCAR, we can reason about our own cognition and decide to redirect its course in various ways. When we do this we are engaging in practical (decision-theoretic) reasoning about how to proceed. Although OSCAR can do this in principle, at its current stage of development OSCAR performs no reflexive cognition. It is this ability to deviate from our default rules of cognition that enables us to behave irrationally. For example, I noted that we can decide what to think about. This makes us more effective problem solvers, but it can also lead to irrational behavior. For instance, every researcher has had the experience of having occur to him a possible difficulty for a favored theory, and felt the temptation to ignore it, i.e., to reorder his cognitive task queue so that he never thinks about it again. That would be irrational, and what makes it possible is reflexive cognition.

In my (2007), I surveyed cases of human irrationality, and argued that they can all be traced to a single source. I remarked that a large proportion of our judgments are the result of applying Q&I

[7] I was one of the first philosophers to write about defeasible reasoning, beginning in my 1965 PhD dissertation, and then in my (1967), (1970), (1971), (1974), (1986). Defeasible reasoning formed the cornerstone of my epistemology in my (1974) and (1986), and its implementation has been one of my main interests in AI — see my (1995) and (2002). See my (2007a) for a sketch of my general theory of defeasible reasoning.

modules. However, rationality dictates that when we have reason for being suspicious of the outcome of a Q&I module, we should override it by reasoning explicitly about the matters at hand. Unfortunately, we often have difficulty overriding Q&I modules. I am unsure whether this difficulty manifests itself in epistemic cognition, but it obviously does in practical cognition. Explicit decision-theoretic reasoning is difficult for us, because we often lack knowledge of the requisite probabilities and utilities. We get around this by employing a wide range of Q&I modules for practical decision making. For example, we take a desire to do something to be a reason for doing it. But sometimes fulfilling a desire can have long-term negative effects. Consider the conditioned desire to smoke cigarettes. On the basis of that desire, many people smoke. Most of them know that smoking is bad for them, and that they should not smoke, but they do it anyway. In other words, their explicit decision-theoretic reasoning does not have the power to override their conditioned desire. I discussed these matters at length in my (2006). Insofar as a person fails to override a Q&I module by explicit reasoning leading to contrary conclusions, he is being irrational. In my (2007), I suggested that this is the sole source of human irrationality.

4. The Role of Rationality in a Theory of Cognition

We can distinguish between two kinds of cognition. Much of our cognition, like the construction of the visual image on the basis of perceptual input, proceeds automatically and we cannot voluntarily alter its course (except indirectly, e.g., by closing our eyes). But some aspects of our cognition are "voluntary", in the sense that we can engage in reflexive reasoning about them and decide whether to follow our default rules, or do something else. Rationality only pertains to the latter. The production of the visual image can be erroneous, in the sense of misrepresenting our surroundings, but it cannot be irrational. But both epistemic and practical cognition can be overridden by reflexive cognition, and as such they can be irrational.

The traditional view of philosophical theories of practical and epistemic cognition is that they are normative. They are about how we *ought to cognize*, and as such they are orthogonal to psychological theories of cognition, which are about how we *do cognize*. But I think this view is wrong. I believe that both philosophical and psychological theories of rationality are empirical theories about the contingent structure of our cognitive architecture. I argued that philosophical theories are based on our philosophical intuitions about how we ought to cognize, but these in turn are a contingent aspect of our cognitive architecture. The theories of rational cognition that we construct on the basis of these philosophical intuitions are theories about the structure of our cognitive architecture. In particular, they are often theories about the default rules that we follow when we are not engaging in any reflexive cognition. However, reflexive cognition itself is not different in kind from other cognition. It is just a matter of turning our general-purpose reasoning procedures on a different subject — itself.

The conclusions we draw in this way about the structure our cognitive architecture could in principle be discovered by the psychologist using non-philosophical methods. We could arrive at

the same theories in either way. However, philosophical theorizing trades upon the fact that our cognitive architecture affords us privileged access to certain features of our cognitive architecture, via our philosophical intuitions. That it does this is just one more contingent feature of that architecture. This privileged access does not apprise us directly of how our cognitive architecture works. Rather, it gives us particular instances of rational or irrational cognition, and we can take that as our data for constructing general theories of cognition. The construction and confirmation of theories on the basis of this data works the same as theory confirmation does anywhere in science. There is nothing uniquely philosophical about it.[8]

The preceding remarks suggest that philosophical theories of rational cognition are perfectly ordinary theories about certain aspects of human cognition, not essentially different from psychological theories. However, it cannot be denied that there is something normative about judgments of rationality. If I become convinced that I have cognized irrationally, that moves me to try to correct my cognition. Judgments of rationality are normative in the sense that they provide assessments of value that plug into practical cognition and effect our subsequent behavior. We value being rational, and that moves us to try to achieve it. However, this is just one more aspect of our cognitive architecture. It builds in loops whereby judgments about rationality affect the subsequent operation of the architecture. This is a sense in which philosophical theories are normative, because they issue in judgments of rationality, but all of this is again a contingent aspect of our cognitive architecture, and it is the sort of thing that psychologists must study just as much as philosophers if they are to obtain a complete theory of human cognition.

The upshot is that philosophical theories of rationality are not different in kind from psychological theories of cognition. They focus on a restricted subclass of cognition — voluntary cognition — and they are based on a different methodology. However, the theories produced are about the contingent structure of the human cognitive architecture. The theories have normative import, because our cognitive architecture involves loops whereby judgments of rationality affect the course of cognition. But this would be no less true if the theories were produced by psychologists rather than philosophers. Ultimately, the study of rational cognition belongs more generally to cognitive science — not just philosophy or psychology. Their research methods are complimentary rather than in competition. And this makes it easier to acknowledge that traditional philosophical methodology, with its emphasis on armchair thought experiments, is too limited even for philosophy. To get theories of rational cognition right, we must test our theories by engaging in computer modeling.

4. Bibliography

Cheng, P. W., and Holyoak, K. J.

1985 "Pragmatic reasoning schemas". *Cognitive Psychology* **17**, 391–406.

[8] These observations derive from my (1986) and (2007).

Pollock, John

1967 "Criteria and our Knowledge of the Material World", *The Philosophical Review*, **76**, 28-60.

1970 "The structure of epistemic justification", *American Philosophical Quarterly*, monograph series 4: 62-78.

1971 "Perceptual Knowledge", *Philosophical Review*, **80**, 287-319.

1974 *Knowledge and Justification*, Princeton University Press.

1986 *Contemporary Theories of Knowledge*, Rowman and Littlefield.

1989 *How to Build a Person*. Bradford/MIT Press.

1995 *Cognitive Carpentry*, MIT Press.

1998 "Procedural epistemology — at the interface of Philosophy and AI", in *The Blackwell Guide to Epistemology*, ed. John Greco and Ernest Sosa, Basil Blackwell.

1998a "Perceiving and reasoning about a changing world", *Computational Intelligence*. **14**, 498-562.

2002 "Defeasible reasoning with variable degrees of justification", *Artificial Intelligence* **133**, 233-282.

2006 *Thinking about Acting: Logical Foundations for Rational Decision Making*, New York: Oxford University Press.

2007 "Irrationality and cognition", in *Epistemology: New Philosophical Essays*, ed. Quentin Smith, New York: Oxford University Press.

2007a "Defeasible reasoning", in *Reasoning: Studies of Human Inference and its Foundations*, (ed) Jonathan Adler and Lance Rips, Cambridge: Cambridge University Press.

Pollock, John, and Joseph Cruz

1999 *Contemporary Theories of Knowledge*, 2nd edition, Lanham, Maryland: Rowman and Littlefield.

Simons, D. J., & Chabris, C. F.

1999 "Gorillas in our midst: Sustained inattentional blindness for dynamic events." *Perception* **28**, 1059-1074.

Wason, P.

1966 "Reasoning". In B. Foss (ed.), *New Horizons in Psychology*. Harmondsworth, England: Penguin.