ROBERT CUMMINS, PIERRE POIRIER and MARTIN ROTH

EPISTEMOLOGICAL STRATA AND THE RULES OF RIGHT REASON

ABSTRACT. It has been commonplace in epistemology since its inception to idealize away from computational resource constraints, i.e., from the constraints of time and memory. One thought is that a kind of ideal rationality can be specified that ignores the constraints imposed by limited time and memory, and that actual cognitive performance can be seen as an interaction between the norms of ideal rationality and the practicalities of time and memory limitations. But a cornerstone of naturalistic epistemology is that normative assessment is constrained by capacities: you cannot require someone to do something they cannot or, as it is usually put, *ought implies can*. This much we take to be uncontroversial. We argue that differences in architectures, goals and resources imply substantial differences in capacity, and that some of these differences are ineliminable. It follows that some differences in goals and architectural and computational resources matter at the normative level: they constrain what principles of normative epistemology can be used to describe and prescribe their behavior. As a result, we can expect there to be important *epistemic* differences between the way brains, individuals, and science work.

1. INTRODUCTION

We have inherited a tradition in Philosophy that distinguishes Epistemology proper, which focuses on knowledge and the evidential justification of belief, from the broader study of rationality, which encompasses practical reason, i.e., rational action, as well as rational belief. There are various ways of collapsing this distinction. One might, for example, hold that what makes a belief rational is not the evidence for it, but rather its tendency to generate rational actions. More conservatively, if one assumes, contra Hume (1739), that adopting a belief is an action, then the theory of rational belief becomes a special case of the theory of rational action. Conversely, one might think that an action is rational just in case it is rational to believe it is the best alternative, in which case the theory of rational action becomes a special case in the theory of rational belief.

Whichever position was taken on the relation between rational belief and rational action, it was generally assumed until recently that rational agents were individual intelligent beings – typically adult humans – that are capable of reflective consciousness and propositional attitudes. The

Synthese **00:** 1–45, 2003.

theory of rationality generally was couched in the language of propositional attitudes and reflective consciousness. The development of cognitive science, however, eventually brought both assumptions into question. Functionalists generally, and AI researchers in particular, assumed that intelligent reasoning was possible without consciousness. Frustrated with traditionalist quibbles about the claim that speakers of a natural language have tacit knowledge of its grammar, Noam Chomsky introduced the verb 'cognize' and its cognates to express the concept he thought was needed for the newly emerging science of psycholinguistics. You get the concept of cognition by, as it were, subtracting from the concept of knowledge everything not relevant to psychological function. It is like belief in that it needn't be true or justified, but unlike belief in that it needn't be even potentially conscious. Many philosophers, indeed, adopted the view that beliefs need not be even potentially conscious (Fodor 1968), in order to accommodate the manifest need to posit unconscious inference (Helmholtz 1866). This liberalization in the theory of cognition opened the door for cognizers other than adult humans. Infant humans, non-human animals, and what Dennett (1969, 1978) called sub-personal agencies, became proper subjects for cognitive states and processes, and hence for normative epistemological assessment. It began to seem plausible – even obvious – that reason and inference were central to the explanation of a great variety of psychological effects and capacities.

Early Cognitive Science, however, still took rationality to be a phenomenon that lived on the propositional attitudes: it was Helmholtzian through and through. Along with the connectionist revolution, however, came the idea that cognition might not require propositional attitudes at all. Eliminativism became a serious position (Churchland 1981, 1989). John Haugeland published a seminal paper (1991) suggesting that non-linguistic representations such as pictures or activation patterns do not have propositional contents at all. It suddenly became clear that there are many kinds of representations – scale models, pictures, graphs, diagrams, maps, synaptic weight configurations and partitioned activation spaces, activation vectors – that can be evaluated for accuracy, often on several simultaneous dimensions, but not for truth: they do not represent propositions, and so they do not have truth conditions. Truth-conditional semantics began to seem inadequate to account for the relations between representation and world.

The idea that cognition might traffic in representations that do not have propositional contents is a far more radical shift in the theory of rationality than the introduction of computationally realized unconscious beliefs, inferences, and sentences in a language of thought. How can there be

rationality if there is not inference, and what is inference if there are no propositional contents, hence no truth? Could logic be irrelevant to reason?

As we come to understand both the brain and science better, it seems clear that we must take very seriously the possibility that a great deal of cognition is not inference as traditionally conceived in terms of propositions and truth. We need to accommodate representations that can be more or less accurate along many dimensions. The rationales that are realized as the disciplined processing of these sorts of representations are not expressible in the idiom of logic or of the propositional attitudes, and that implies that traditional epistemology has little to tell us about the Rules of Right Reason as these are actually found in nature and culture.[1]

It is our contention that the Rules of Right Reason are diverse. This contention stems not from cultural or semantic relativism, but rather from recognition of the vast diversity to be found in cognitive systems. They differ in architecture, in computational and representational resources, and in their functions and goals. All of these differences, we will argue, make a difference. They confront us with a kind of pluralism in the study of rationality and cognitive systems undreamt of by the relativists.

What's left of the concept of a cognitive system when we strip away the focus on consciousness, the propositional attitudes, and even truth? It can have little to do with justified true belief or with inference as conceived in logic. What, then, makes a system cognitive? We have no very satisfying answer to this question. Indeed, our very commitment to cognitive diversity makes us suspicious of the idea that there is a single essence of cognition. That idea was plausible as long as the subject was understood to be the conscious deliberations of adult humans about the true and the evident. It was even plausible when the subject was the computationally realized manipulations of truth-evaluable states (attitudes) or representations. It is *not* very plausible in the current state of neuroscience or the philosophy of science (Churchland 1989; Galison 1987; Giere 1988; Hutchins 1995a; Teller 2001). Our growing understanding of the brain and of the nature of scientific inquiry makes it very unlikely that the representations and rationales of either the brain or of science can be understood in traditional terms. Nevertheless, we think there is a cluster of features that seem to characterize the kinds of systems we think a future epistemology and cognitive science should target:

- Cognitive systems are information driven. To understand a cognitive system, you need to appreciate what information (and how much) it has about itself and its environment and/or containing systems.

- Some of the information is represented, i.e., encoded in objects or processes or states from which structural features of the things represented can be systematically recovered.[2]
- Cognitive systems are relatively complex, with the interactions among sub-systems and components being themselves information driven.
- They are intelligent.
- They are relatively plastic – i.e., they can learn, either individually, or as a type (species).
- They are goal directed, hence characterized by functions that are at least implicitly normative. They can, in short, be rational or irrational.

Not every cognitive system has all of these features, and many, perhaps all, of these features are to be found in non-cognitive systems, though not all together. For us, the concept of a cognitive system is what Putnam (1962) called a cluster concept.

Our thesis is that the Rules of Right Reason are as diverse as the cognitive systems to which they apply. Traditionally, epistemology has been carried out in the framework of the propositional attitudes. While some have deplored this (notably Churchland 1979, 1989), it remains the case that knowledge and belief dominate the scene, with desire and intention playing a role in accounts designed to deal with practical reason. This poses a difficulty for us, since we want to consider the epistemology of institutions like science and of neurally implemented psychological systems such as visual object recognition, neither of which, we hold, are naturally characterized in terms of the propositional attitudes. We thus require a conception of rationality and reason that transcends the limits set by the propositional attitudes. We want to be able to ask whether science, or object recognition, proceed rationally, and, if so, how, without presupposing that science, or object recognition systems, have beliefs or any other propositional attitudes.

We therefore propose to think of our subject matter as rationality rather than knowledge, and to think of rationality in the broadest terms as epistemic constraint satisfaction. This is, regrettably, but inevitably, uninformative, since it is part of our brief that the relevant epistemic constraints differ substantially in kind across epistemic strata. We are confident, however, that there is an epistemological sense of getting things right or wrong, of doing better or worse, that allows us to ask about the rationality of very diverse cognitive systems. The proof of the pudding, as usual, will have to be in the tasting.

2. STRATIFIED EPISTEMOLOGICAL PLURALISM

For reasons we shall not dwell on, philosophers are hypersensitive to the issue of cultural relativism and perhaps that's why they, and cognitive scientists with them, have resisted the sort of epistemological pluralism we intend to defend. We will not address the issue of cultural relativism here other than to say that it is simply a mistake to extend whatever good argument one may mount against it to the whole cognitive or epistemological sphere. To see this, note that epistemological monism runs in two orthogonal dimensions, horizontal and vertical as it were.[3] There is no such thing as epistemological monism (or pluralism) tout court. To get a grip on the distinction, consider a few systems involved in the "cognition business" generally but that nevertheless differ in goals, resources, or architectures (Figure 1). Horizontal epistemological monism is the thesis that there is only one set of epistemological norms appropriate to *any one level in the schema*. It is the claim, for instance, that all cultures may be assessed by reference to a single set of epistemological norms, or that all individuals (agents, scientists) may be assessed by reference to a single set of norms. Vertical epistemological monism, on the other hand, is the thesis that a single set of epistemological norms suffices for the assessment of systems at different levels. It is the thesis that, whatever else it may represent, the figure above does not represent differences in epistemological type in the vertical direction.

Here is a slightly different way to see the difference between the two dimensions. Let's say that pluralism is true in both dimensions. Then the set of all possible sets of norms, and hence the set of all possible types of cognitive system, forms a plane (2-D space). Any point in that space may represent a kind of cognitive system subject to its own unique and proprietary norms. Deny one or the other form of epistemological pluralism and the plane collapses to a line (1-D space). This might be the case if, for instance, the stratified picture of epistemology depicted above turns out to be right but, at any given stratum, a single set of norms suffices for the assessment for every system in that stratum. Deny the last remaining form of pluralism and the line collapses to a point, giving rise to the strongest form of epistemological and cognitive monism: the rules of right reason are one and apply universally.

Given the distinction between horizontal and vertical epistemological pluralism, it is clear that arguments against cultural relativism have no impact on the issues that concern us here. Although the previous figure is nicely suggestive of the epistemic ordering of cognitive systems we propose, proponents of (vertical) epistemological monism will rightly

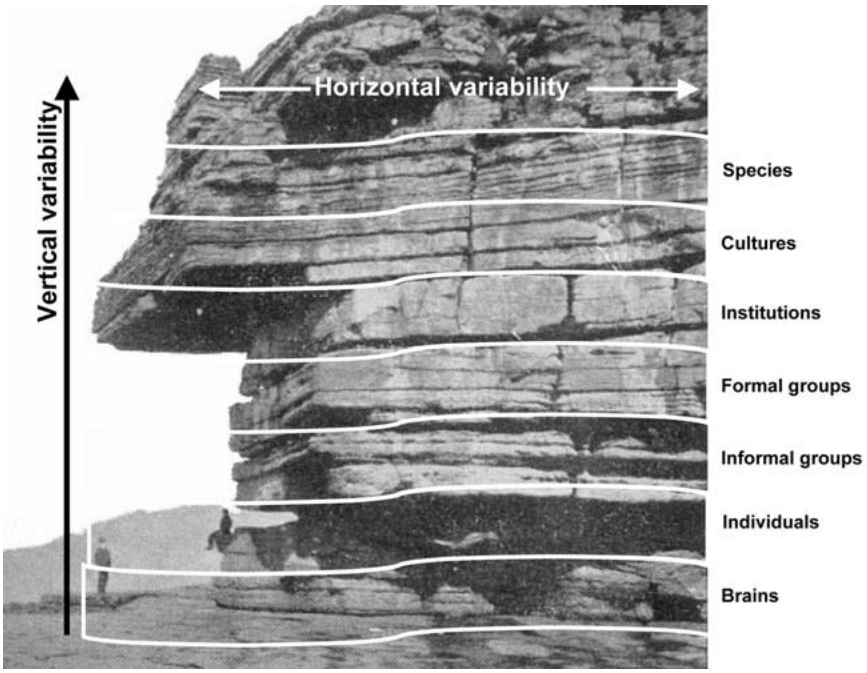

Figure 1. Epistemological strata.

point out that it begs the question against their view. Given what we said about the prevalence of epistemological monism in epistemology and cognitive science, surely we don't get to stipulate, e.g., that brains and formal institutions belong to different epistemological strata. That would be *Philosophy Made Simple*; much too simple. In this section, we will argue that epistemic strata typically differ with respect to their computational architecture, their representational resources, their computational resources, and their function or goal. These differences entrain different epistemic constraints. Hence, the 'rules' or norms of epistemic assessment, and the concepts of epistemic description, differ substantially across strata. The rejection of (vertical) epistemological monism follows.

We believe the failure to recognize the existence of vertical pluralism has given credence to theories that fundamentally mischaracterize their object and that horizontal monism at the brain level still blinds both scientists and philosophers to the (horizontal) cognitive diversity they might find in brains. Vertical monism, for instance, has led philosophers and cognitive scientists to believe theories like the following may be true, or useful approximations:

- *Cognitive development as scientific progress*: According to Gopnik and Meltzoff (1997), children's theories are abstract, integrated, and

domain specific. They appeal to causality and make ontological commitment. They function to interpret and explain past or current events and to predict future events. Finally, children replace theories that do not correctly serve their function by theories that do, or do it better, once these become available. In short, children's knowledge is structurally, functionally, and dynamically similar to scientific knowledge. Scientists are grown ups who have kept on improving the theories they developed as children!

- *Language learning as hypothesis confirmation*: According to Fodor (1975) linguistic expressions acquire their meaning by the following general process. Hypotheses about the possible meaning of expressions (words, sentences) are formulated and then checked against incoming empirical evidence. Confirmed hypotheses are kept and come to constitute the individual's theory of meaning. Hence, the acquisition of meaning (at the linguistic, not LOT, level) is the outcome of a processes of hypothesis formation and confirmation. We acquire the meaning of words by a systematic application of the scientific method![4]
- *Scientific progress as movement in weight space*: According to Paul Churchland's (1989) neurocomputational perspective on epistemology, a new scientific theory is acquired like this: an internal trainer "sees" that the predictions made by the current theories do not fit the incoming evidence, computes the gap between what the theory predicted and what it should have predicted (incoming evidence), determines to what extent each of the billions of synaptic connections involved in the processing contributed to the gap and then adjusts the strength of their connections accordingly. A scientific theory is a matrix of synaptic connections!

We would like to draw attention to a basic assumption shared by all of these theories: that scientific, individual, and neural cognition are fundamentally the same kind of process, and can be epistemologically described and evaluated by reference to the same norms.[5] The idea that the Rules of Right Reason are everywhere the same is encouraged by the further assumption that there is really only one kind of cognitive system. This is what makes it natural to use models developed by philosophers of science to describe cognitive processes or to use models developed by neurocomputational scientists in order to describe scientific processes.

We shall argue that the previous figure indeed depicts an *epistemic* ordering of cognitive systems. The depicted ordering may not be right in all details: some may contend that a proposed level is not really different from another, or that we have forgotten an important level. We are concerned

here simply to make a case for the idea that epistemology comes organized in this geological manner, with different kinds of diversity appearing within strata and across strata. In a nutshell, the argument goes like this. A cornerstone of naturalistic epistemology is that normative assessment is constrained by capacities: you cannot require someone to do something they cannot or, as it is usually put, *ought implies can*. This much we take to be uncontroversial. But we shall show that different architectures, goals and resources imply substantial differences in capacity, and that some of these differences are ineliminable. It follows that some architectural, goal and resource differences matter at the normative level: they constrain what principles normative epistemology – what Rules of Right Reason – can be used to describe and prescribe their behavior.

Systems generally can be ordered in various vertical hierarchies: constituency, functional, supervenience, etc. Although we shall not argue for it here (we reserve that issue for later), we believe that epistemic systems can be vertically ordered in terms of their respective architectures. The relevant vertical hierarchy for architectures seems to be a partial constituency hierarchy, that is, a hierarchy ordered by partial constituency, or part-whole (mereological) relations. Architectures at one level are components of architectures at the levels above. If, epistemologically, the architectures that belong to this hierarchy are essentially diverse, that is they cannot be reduced to a general architecture affording similar capacities to systems, it follows from this and the thesis that some architectural differences matter normatively that there is a vertical epistemic ordering of systems. Read in this way, the previous figure says for instance (1) that individuals and brains are both full fledged epistemic systems, (2) that they are *different* epistemic systems, and (3) that brain architecture is a constituent of individual organism architecture; likewise, individuals and institutions are both full fledged yet distinct epistemic systems where the former are components of the latter; and so on.[6] We shall call each level in this vertical epistemic hierarchy, an *epistemological stratum*.[7] Science, individuals and brains (but also perhaps governments, companies and courts of law (Goldman 1999)) are all cognitive systems in their own right. They are systems whose main business is to satisfy some set of epistemic constraints (e.g., innocents should not be convicted). It follows that to properly explain a system's cognitive capacities, you must refer to the stratum which defines the appropriate norm. That is why we think, for example, that science is not a good model of individual cognition, and that individual cognition is not a good model of science.

3. COMPUTATIONAL RESOURCES

It has been commonplace in epistemology since its inception to idealize away from computational resource constraints, i.e., from the constraints of time and memory. One thought is that a kind of ideal rationality can be specified that ignores the constraints imposed by limited time and memory, and that actual cognitive performance can be seen as an interaction between the norms of ideal rationality and the practicalities of time and memory limitations. This amounts to a competence-performance distinction of the sort familiar in linguistics (Chomsky 1965). The underlying assumption is that normal humans actually are ideal reasoners, but that their performance is less than ideal because of the influence of resource constraints, failures of attention or motivation, insufficient or inaccurate information, and the like.[8] Another approach is to assume that failures to satisfy 'ideal' norms of rationality are simply the result of ignorance or stupidity: some people are simply more rational than others. An increasingly common response to apparently irrational judgments and decisions, however, is to argue that the behavior in question is not in fact irrational, but the normal performance of a cognitive system operating within the constraints imposed by real time and space. In this vein, it has been argued that some apparently irrational behavior is the expected outcome of pursuing a strategy designed to deal with specific resource limitations.

Because 'ought' implies 'can', the norms of rationality cannot require a cognitive system to do what it cannot do. Hence, we should not apply a norm to a case in which it cannot be satisfied. More generally, we should not apply a norm that requires abandonment of a strategy that has no feasible alternative that performs better. Thus, the normative theory of rationality must be a theory that evaluates performance, practice, strategies, etc., in the light of what is possible/likely for the system being evaluated. What a system can do evidently depends on its computational resources, i.e., on its time and memory. A cognitive system that is part of a system in either the perception or action business – i.e., whose goals are *practical* – is continually faced with time constraints. And because all cognitive systems are finite, memory is always limited in various ways.[9]

It is our contention that a normative theory of rationality should not idealize away from resource constraints. To idealize in the sense relevant here is to construct models in a way that ignores some factors known to be relevant in order to achieve a core theory about the influence of other factors thought to be more basic. Thus, the ideal pendulum law models the period of a pendulum as a function of length and gravitational acceleration, idealizing away from such things as friction and air resistance. This is

possible because there is such a thing as the way the pendulum would behave were there no friction or air resistance. We cannot idealize away from length or gravity in the same way, because there is no way the pendulum would behave (as a pendulum) were there no gravity, or were it to have zero length. Traditional epistemology has treated resource constraints as analogous to friction and air resistance in mechanics. But there are strong reasons for thinking that this is a mistake: Effective cognitive strategies are typically designed to work within quite specific resource constraints, and work poorly or not at all when these constraints are altered. We can get a handle on pendulum behavior by asking how friction and air resistance would modify the behavior of an 'ideal' pendulum. But a reasoning system that would work with unlimited time and/or memory will not typically work *imperfectl*y with limited resources, in the way that, say, a bubble algorithm for sorting a list will; reasoning systems given unlimited time and memory *will generally not work at all*. Part of the issue here is that genuine real world cognitive problems often have time and memory constraints built in: The problem is to do something or other – e.g., object recognition – in 500 milliseconds, or using seven chunks of short term memory. Idealizing away from memory or time constraints in cases like these simply makes no sense. Equally important, however, is the fact that many cognitive algorithms are provably ineffective or non-terminating except under rather specific time and memory constraints. Both of these points will be illustrated in the next two sections. We believe that the implication of these cases is clear: ideal agents as traditionally conceived may not be idealizations of any actual agent, and hence "ideal agent" epistemology may give us little or no insight into genuine rationality.

3.1. *Time*

A predator recognition system is accurate to the extent that it identifies all approaching predators, and avoids misidentifying non-predators as predators. A system behaving so as to maximize accuracy, however, is not necessarily behaving rationally. This is because the point of predator recognition is (typically, anyway) predator avoidance, so speed is of the essence. Natural predator recognition is therefore hair-triggered: in the interest of speed, many false positives are tolerated, while false negatives are kept to a minimum. A rational predator recognition system will sacrifice accuracy for speed. This much is widely recognized. What is not so widely appreciated is the fact that such a system must be designed very differently – be based on very different principles – than one that maximizes accuracy. In particular, one cannot hope to build a rational predator recognition system by building an accurate one and then subjecting it to real world

time constraints. An organism harboring such a system will not survive to reproduce.

If we are to uncover the principles of rational predator recognition, then, we must incorporate the temporal constraints from the very beginning, for they are part of the essence of the problem. The Rules of Right Reason, in this case, depend essentially on how much time the prey typically has to escape in its typical environment from its typical prey. This will depend, of course, on the capabilities of the predator in the environment in question, and on the capabilities of the prey in that environment. Amphibians must behave differently on land than in the water; snow changes everything for land animals, and darkness changes everything for almost every organism. The time frame also depends on supporting behaviors, such as how far a prairie dog wanders from the nearest hole. More daring forays require earlier warnings. Lookouts can devote more resources to looking out, freeing foragers to devote more resources to foraging. A recognition system that lacks variable sensitivity may be rational in an animal that engages in supporting behaviors that keep the average time from warning to safety within rather narrow boundaries, but irrational in an animal that lacks such supporting behaviors.

These sorts of points are now widely conceded for non-human cognition and even for human cognition at the sub-personal level (Fodor 1983, 2000; Pollock 1989; Pinker 1997). But they are not widely thought to apply to the targets of traditional epistemology, viz., science and adult human reasoning. It is certainly no part of our brief to claim that temporal constraints are intrinsic to the design of every cognitive system. Indeed, the argument for essential diversity in cognition is strengthened by noting that temporal constraints are intrinsic to some cognitive designs but not others.

An important example of a cognitive problem that cannot be well-handled by a system designed around fixed time constraints is planning. While planning must take deadlines into account, it is clear that the principles of planning will not themselves vary with the time frame. Rather, the time frame will be, in so far as it is known, an input to the planning process. Planners must often choose among several different planning strategies, because different strategies may be appropriate not only for different problems, but for a given type of problem in different time frames. This choice may be difficult. Sometimes there are no estimates for how long a strategy can be expected to take; sometimes there are (walking vs. driving). Estimating time itself takes time and memory, as does the scheduling for which these estimates are inputs. Different sorts of estimates may be appropriate in different circumstances. Worst case estimates may be appropriate when the price of error is very high; average case estimates may be appropriate

when the problem is one often encountered and what matters is average performance. Moreover, deadlines come in different flavors. In addition to simple deadlines (respond by t), there are contingent deadlines, i.e., cases requiring a response before a certain event takes place, and these are complicated by the fact that predicting when that event will occur may be more difficult early in the planning process than later. Finally, the amount of time available for planning and acting may depend not only on deadlines, but on what other tasks must be done during the same period, and how long these will take. Often, it is not known what other tasks will have to be done, nor how long they will take. Acquiring information concerning likely competing tasks itself takes time. And so on.

It is precisely because temporal considerations tend to be moving targets in planning that planners cannot typically be designed to operate within a specific time frame. Important as this point is, however, it should not be allowed to obscure the fact that many cognitive systems are not planners, and are designed to operate within quite specific time constraints. For instance, many complex cognitive systems need to synchronize processing: output from one subsystem needs to get to other subsystems at just the right time, not earlier, not later. Fluid motor control is one such case. It cannot rest on feedback from perception of the environment since there is no way the signal feeding back from the environment can get to the systems that control movement in time. Instead, those systems must rely on signals from an "emulator" whose function it is to predict the feedback signal, and that prediction must get to motor controllers just when they need it, that is before the actual feedback signal comes in. That signal, when it does come in, is then used to fine-tune the emulator's predictions (see Clark 1995). The production and comprehension of speech are also fast processes that require precise timing. If a system that provides input to these processes is too slow, the proper sequencing of speech can be affected and conditions such as dyslexia may result. It has been shown that in some dyslexics, the magnocellular pathways, which transmit rapidly changing visual sensations, are slower than in normal subjects (Galaburda and Livingstone 1993). For these systems, an epistemology that idealizes away from time constraints will simply misrepresent the problem. An understandable tendency to focus on propositional knowledge or on human planning obscures the relation temporal constraints often have to the substance of Right Reason.

While it is widely appreciated that most cognitive processes will fail if they are not given enough time, it is, perhaps, less widely appreciated that too much time can doom a process as surely as too little. Almost everyone who has ever taken a standardized test has been advised to "go with

your first impression". This is excellent advice for answering questions involving a substantial component of perceptual recognition in which too much "thinking" is more likely to introduce intrusions than to uncover relevant information. Many neural network models that involve recurrent activation give optimal output within a rather narrow time window. Outside that window, on either side, output degrades, sometimes quite sharply. (See for instance, the Jets and Sharks Network, McClelland and Rumelhart 1988.)

There is another way in which the timing of cognition can be critical. Perception (and certain other information gathering processes) must often be geared to the speed with which events transpire in the information source. A perceptual system that works too slowly will simply not keep up with the distal process it is supposed to monitor or sample. Many language-impaired children (SLI and dyslexia), for example, simply process acoustic information too slowly to distinguish very rapid frequency changes and their capacities significantly improve when the speech stimulus is slowed down, for instance by using computer-generated synthetic speech (see e.g., Tallal et al. 1997). Processing information too slowly for real-world events also explains why the famous "keep your eyes on the ball" strategy just won't work at bat: tracking a pitch in baseball is a hopeless strategy (D. Cummins 1995).

More subtly, the rate at which the source is sampled can have a profound effect on what kind of information is gathered. Speed up an audio recording of a Bach partita and elements of structure previously difficult to discern will pop out at you. The structure of a philosophical argument may be lost to the reader who reads too slowly and carefully. When it comes to information, more is not always better, and higher-level patterns can sometimes be made to appear by optimizing the speed of the reasoner relative to a source that is not under its control.

3.2. *Memory*

Cognitive systems differ widely in their memory resources. Not only do cognitive systems differ in the amount of memory available, they differ in the way memory is organized (e.g., is there a long-/short-term distinction?), how it is accessed (random vs. serial; constructive vs. literal retrieval), the kinds of things that can be stored (symbols, pictures, models, samples, simulations, etc.), and the nature of the information (e.g., semantic vs. episodic memory). All of these differences make a difference to the sorts of task a system can do, or do easily. A scientist can often store an actual sample of material (bone, virus, DNA), whereas a brain must store a representation, or something from which a representation can

be constructed. Scientific method need not be specifically designed to deal with the sort of limitations of short-term memory that confront organisms, or with the in-principle limitations of access that characterize the modular architecture of the brain. More and better information is surely beneficial to science, but can be harmful for systems with limited or unreliable memory, or slow and unreliable access to it. There are important interactions with time, here. A large, loosely organized store of information is a liability to a system operating under serious time constraints, but may be advantageous to systems such as science in which time is not a factor and prejudging issues of classification and relevance is dangerous.

Computational modeling has shown that specific limitations on the size of short-term memory can facilitate the solution of certain problems. Increasing or decreasing the size of short-term memory can make whole classes of problems difficult or impossible. It is known that short-term memory increases until adulthood (Dempser 1981) and some (Newport 1990; Goldowsky and Newport 1990) have argued that learning a first language may be an insoluble problem for anyone with *too much* short-term memory, that is, adults. Tasks such as the acquisition of morphology, which involve componential analyses and give rise to a large number of computations to be performed, may benefit from a limited short-term memory: the loss of possibly relevant data, which slows down the acquisition of morphology, is greatly compensated by the reduced computational strain, which makes morphology learnable in the first place (see also Elman 1993 for a connectionist simulation of the interaction between size constraints on short-term memory and language learnability). These results make it clear that the Rules of Right Reason must often be specifically tuned to the size of short-term memory. Natural and artificial design will tend to converge on short-term memory resources that favor solutions to those classes of problems whose solutions are important to survival and replication.

3.3. *Time-Memory Trade-Offs*

A rule of thumb in computer science is that there is a time-memory trade-off: you can do with less memory if you have more time, and less time if you have more memory. While there is no doubt a significant kernel of truth here, it is a truth largely limited to considerations of computability: a function may be computable in less time if there is more memory and vice versa. It is important to realize, however, that the algorithm required in the free-time limited-memory case will typically be very different than the one required in the big-memory but limited-time scenario. Thus, the point about time and memory being exchangeable should not lead us to think that a given cognitive system is typically in a position to exchange

them, and hence that computational resource considerations are extrinsic to the principles involved in the computation.

Real cognitive systems are seldom simply in the function-computing business in the sense in which this means simply paring values with arguments. They are physical systems embedded in the real world, faced with real world problems. Right Reason, in these circumstances, requires getting along with the resources one has. The crucial thing to keep in mind when considering such systems is not the trade-off, but the interaction: a system's memory must be geared to its time constraints, and vice versa.

* * *

In the forgoing, we have illustrated the idea that the substantive content of rationality varies with resources. It is our belief that this situation is ubiquitous: cognitive systems generally are designed to fit their computational resources. This is no less true of institutions such as research science than it is of psychological modules such as the human face recognition system. Strategies that are effective in the context of one set of computational resource constraints are unlikely to be effective if that context is substantially altered. It follows that the substantive content of Right Reason must often be different for systems that operate with different resources. Putting a scientific zoological or linguistic theory in the head is unlikely to yield rational predator recognition or rational speech processing. Conversely, rational predator recognition or speech processing strategies are unlikely to make good zoology or linguistics.

Idealizing away from resource constraints not only makes for bad cognitive science, it makes for bad normative epistemology. In particular it is generally a mistake to ground epistemological assessments or prescriptions on counterfactuals concerning what would have happened had the system taken more (or less) time, or remembered some fact or principle relevant to the "correct" solution. The mistake is two-fold. First, achieving the correct solution is not always or even typically the same as behaving rationally. Second, what would have happened had the system operated outside of its normal resource constraints is typically some sort of crash or disastrously maladaptive response. While our examples are limited to a few strata, we believe the point generalizes, or more circumspectly, that it shifts the burden to those who favor idealization to show that it is legitimate in the stratum in question. This point should be kept in mind when assessing the argument of this section and the next.

4. REPRESENTATIONAL RESOURCES

Consider a map making concern whose function is to represent the street and intersection structure of various cities on paper road maps. For such a system, maximizing accuracy is not the same as maximizing effectiveness. Accuracy, beyond a certain point is often expensive in terms of computational resources, and may sometimes lead to intractable representations. A city map that shows all the streets will often be too big to read, unless the reduction is too small for anyone to see. A standard compromise in road atlases is to show only main routes. In general, an effective representing mechanism, therefore, will produce representations that are accurate enough to enable its "customers" to get their jobs done, while still tractable, timely, and not too memory intensive. For a similar reason, NASA does not use relativistic mechanics when performing the necessary calculations to send a space shuttle into space since using Newtonian mechanics is much simpler and accurate enough to get the job done. NASA thus tolerates some representational inaccuracy in order to ensure easy computational tractability. Trading some representational accuracy in order to insure an important property is obtained is also something the brain does. Retinal neurons, and other neurons farther down the visual pathways, tend to exaggerate differences between dark and light. As a consequence, fuzzy boundaries become much sharper. The purpose of this, of course, is to enable consumers down stream to detect the edges and borders of things. These cells sacrifice accuracy with respect to the actual contrast in order to facilitate detection of ecologically important properties like edges.

The foregoing is enough to make it clear that it can often be a rational strategy to sacrifice accuracy for effectiveness. Two further points, however, are worth emphasizing. First, the relation of effectiveness to accuracy depends on when accuracy is really important. An effective representer will operate so as to generate relatively accurate representations when accuracy is important – i.e., when representational error is potentially serious. Hence, a rationally designed representer may embody principles that make for accuracy in a special subclass of cases in which accuracy matters most, with the consequence that targets differing substantially from members of the favored subclass are systematically represented less accurately. Second, since many kinds of representations can be assessed for accuracy on more than one dimension, a rational representer will embody a design that favors accuracy along those dimensions in which error is serious, with the (often inevitable) consequence that relatively large but less serious errors along other dimensions are common. Mercator projection maps are a good example here. They are accurate with respect to true direction, and

thus valuable to navigators. However, as land masses get more distant from the equator, representations of their shape and size get quite distorted. Scientists, too, often trade accuracy in one dimension in favor of accuracy in another. When explaining the diffusion of some substance in water, scientists think of water as a collection of molecules. When explaining the flow of water in a pipe, scientists think of water as a continuous, incompressible medium. Each of these models of water is accurate in some respects and wildly inaccurate in others. Each ignores or distorts some properties of water in order to accurately represent others, depending on the purpose of the investigator.[10]

Particularly important to the assessment of rationality is the phenomenon of forced error (Cummins 1996), i.e., limitations on representational accuracy that result from principled limitations on the representational capacities of the system. Flat maps of the globe must distort something, and verbal descriptions of things such as faces must leave out a wealth of detail available in even a poor photograph or sketch. Effective representation will minimize serious errors, while tolerating even rather large errors that are not serious given the needs of the consumer. Thus, assessing the rationality of a representing mechanism – what Cummins (1996) calls an intender – requires attention to the needs of the systems consuming the representations. And since a given intender may have multiple clients with different needs, its rationality must be assessed in the light of the perhaps conflicting demands it faces.

Different cognitive systems come equipped with different representational resources: Science, educated common sense, and various neural systems utilize markedly different representational schemes in the service of different goals, and subject to vastly different computational constraints. Systems that differ substantially in representational resources will have fundamentally different cognitive capacities. Epistemological pluralism follows from the observation that the Rules of Right Reason must differ for systems that differ substantially in their cognitive capacities.

5. SUCCESS AND EFFECTIVENESS

In Section 4 above, we distinguished between the accuracy of a representation and its effectiveness, the core idea being that accuracy is often expensive to produce and intractable (indigestible?) to consumers, and hence not always effective. Here, we want to generalize that distinction in order to make a point about the performance of cognitive systems whose goal is not the production of a representation.

Some systems are cognitive in their own right, as it were, and others are cognitive because their behavior is a function, in part, of a cognitive component. Thus, Helmholtz (1866) held that the visual system is cognitive because its performance depends on unconscious inferences. Early pioneers of the cognitive revolution extended this idea to every aspect of intelligent behavior (Cummins and Cummins 2000; Cummins 1983). The actions of human agents, while certainly assessable for rationality, are not, typically, semantically individuated behaviors that can be assessed for accuracy–hence the traditional distinction between practical and theoretical reason. A similar point holds for much institutional behavior as well, and perhaps for many aspects of brain functioning.

With some trepidation, we will use 'degree of success' to denote how well a system achieves its performance goal, and 'rate of success' to denote the rate at which a system hits its performance targets with some specified degree of success. Our contention is that, just as a representation's accuracy must be distinguished from its effectiveness, so a system's success must be distinguished from its effectiveness.

5.1. *Degree of Success (in which Granny takes a hit)*

Granny used to say, "If it is worth doing at all, it is worth doing well". We all know it isn't true. Many things are worth doing, but not worth doing well. By now, the reason should be familiar. Doing things well is computationally and metabolically expensive, and often generates a level of detail or precision or sheer volume that is intractable for consumers. Governments do this as a matter of deplorable routine. This is as true of behavior generally as it is of representation. A perfectly pitched tent is no great benefit if it is ready only after the rain begins. In the movie *Apollo 13* (and, we suppose, in the real mission as well), the crew had to build an air filter out of a limited number of parts and in a limited amount of time. What they ended up with was a crude, poorly effective filter (a design NASA surely would certainly not include in future missions). But success, constrained by time and limited resources, was paramount (the very survival of the crew depended on it) and was bought at the price of temporary low effectiveness. Or take the Mac–PC war. Most agree that Apple makes the better computer, but still the company nearly went bankrupt. The problem, and the idea goes back to Henry Ford, is that if your goal is to sell computers, the rational strategy is not to build the best computer around, but to build one that's good enough for most users but also cheap enough that they can afford it. The same is true, of course, of cars (that's why Ford, whose problems with quality are legendary, now owns Jaguar and other companies that built the better cars – a fate that may yet be Apple's).

5.2. *Rate of Success*

We need to distinguish the single case – how successful the system is on some particular occasion – from the general case – how successful the system is over a population of cases or a period of time. The latter is rate of success.

There are at least three reasons why a system with a high rate of success may be less effective than one with a lower rate of success. First, (predictably) is the fact that a high rate of success can be expensive in terms of resources. Second, some cases may be more important to get right than others, so that a simple percent correct scoring may not be the appropriate measure.[11] Third, systems embedded in others must generally be judged in terms of the contribution they make to the embedding system. This can be extremely complex. Just as the heart is part of both the circulatory and endocrine systems, a single cognitive system may be a component of more than one embedding system. Since every client does not have the same requirements, trade-offs can arise. Avoiding dangerous situations in a timely fashion may require tolerating a high rate of false positives, as we've seen, whereas seeking out beneficial situations may require tolerating a high rate of false negatives. An object recognition system called on to recognize predators as well as prey or potential mates is faced with a difficult balancing act. Failure to take this into account may result in under-rating object recognition generally, or some particular aspect. One reason there are complex and inefficient courtship rituals has to do with the fact that the recognition systems involved need help to compensate for the trade-offs required to satisfy other clients such as predator/enemy detection. There are areas around the sylvian fissure that function in sensory sequencing for both speech and facial mimicking tasks (see Calvin and Ojemann 1994). In humans, we may suppose that the visual system contributes not only to predator recognition but also aesthetic criticism. While a slow but detailed visual system may be a disaster for predator recognition, it may be beneficial for artistic appreciation.

A related point concerns what might be called side effects. Efficient travel is not good exercise. Efficient search tends to minimize serendipitous discovery. Imagine two beavers. Eager Beaver builds dams quickly and efficiently, but they are less stable and leak more than those built by her conservative cousin Careful. Careful builds better dams. But Careful is not necessarily the more rational dam builder. The different strategies imagined for Careful and Eager would typically have many side effects unrelated to dam-building. Movement, foraging and eating patterns will be affected, with their consequent effects on metabolism, encounters with other organisms, including predators and potential mates, and with the

elements generally. The rationality of a cognitive design must be assessed in the light of the impact of its side-effects, as well as in the light of the inevitable trade-offs involved in serving several masters.

Evidently, high rate of success may be possible only at the price of resource consumption, compromising the relatively rare but really important cases, inability to serve more than one client effectively, or side-effects that compromise the containing system, even though they do not compromise the success of the contained system itself.[12]

5.3. *The Norms of Success*

The distinction between effectiveness and success will collapse unless we pay careful attention to how we specify the norm against which success is measured. In particular, if we think of the function or goal of a system as effective performance, then the rate of success will simply be the rate at which the system performs effectively, and the degree of success will simply be the degree of effectiveness. We need to be able to distinguish how often and how accurately a system recognizes predators from how effective that pattern of performance happens to be in some particular context. Selectionist theories of functions that identify the function of a trait with the effect of that trait's incorporation that accounts for its selection will mislead us here, for it is effectiveness that is adaptive, not what we have been calling success. This does not, of course, show that selectionist theories of function are mistaken, but it does show that they do not provide us with a handle on the relevant norms of success. These are, as it were, internal to the system, whereas effectiveness is always measured against an external norm, i.e., a norm that is grounded in the functions or goals of containing system or consumer.

5.4. *Effectiveness and Rationality*

We have been at some pains to distinguish Success from Effectiveness. The tacit implication is that the rationality of a cognitive system is a function of its effectiveness, rather than of its success.[13] While we think this is a step in the right direction, it is too crude because effectiveness can be completely unpredictable. The effectiveness of Careful's dam building strategy as compared to Eager's will depend on how important speed and efficiency are as compared to tightness and stability. That, in turn, will depend on a variety of environmental conditions, some quite unpredictable. A rash of violent floods that sweep away even Careful's marvelous constructions will favor Eager's strategy. Less violent floods will favor Careful's strategy. Thus, how effective a strategy is, and consequently how adaptive it is, can be determined by events the system could not possibly predict. Its

rationality, therefore, needs to be judged in terms of something other than adaptiveness, such as expected effectiveness, i.e., the effectiveness of the strategy over the period and circumstances that shaped its development. This is analogous to traditional wisdom, which holds that rationality is a function of the available evidence, not of the truth.

* * *

This discussion of success and effectiveness presupposes that rationality is a matter of the general case. We have been assuming that we should assess rationality generally in the way rule-utilitarians assess moral action, i.e., in terms of whether it is the manifestation of a rational general strategy. There is a general argument to be made for this assumption. It is evidently not rational for a cognitive system to handle every situation individually. Or rather: treating every situation independently is itself a general strategy that could only be effective in a system with unlimited resources. Real systems need general strategies specialized for certain classes of problems. When the particular departs significantly from the general, the system can be expected to deal with this only if it has some general mechanism for recognizing such exceptions and dealing with them is a special way. But then that mechanism is part of the general case strategy employed by the system, and should be judged accordingly. This will involve such considerations as how expensive is it to be constantly on the look-out for exceptions, how accurately they are identified, how successfully are they dealt with when identified, and the consequences of misidentification.[14]

Once again, we are led to the conclusion that the contents of the Rules of Right Reason are bound to be diverse. Being effective may depend on unpredictable factors. What it takes to be effective will depend on how a system is embedded in its world. A cognitive system must often accommodate many clients; its activities are bound to have side effects; some errors are bound to be more serious than others. These considerations all render implausible in the extreme such suggestions as that the brain works by deploying theories of the sort that science seeks to generate, or that science works by deploying the strategies that make for good object recognition.

6. ARCHITECTURE

Many architectural differences are relevant to our concern about vertical monism. We are not interested here in individual architectures but general classes of architectures. Our aim is to show that the very character that

defines class membership turns on a difference that is relevant to assessing the rationality of the processes implemented by the architectures, and that it is impossible to idealize away from these differences.

The first architectural difference is the sort Fodor and Pylyshyn (1988) refer to in 'Connectionism and Cognitive Architecture', that is, the difference between classical and connectionist architecture. One way to read the paper has Fodor and Pylyshyn claiming that neural networks are simply not cognitive systems at all (except as mere implementations of classical systems), because, according to them, connectionist models cannot account for the systematicity of thought, etc. But that seems strong and, not wanting to be imperialists about what the term "cognitive system" means, they may simply claim that connectionist systems are simply not the right *sort* of cognitive system to exhibit thought and other "central system" or domain general processes (thought, reasoning and the like). Input systems (Fodor 1983) are not productive or systematic,[15] that is the source of their speed, and yet Fodor would surely not claim that they are not cognitive, in some broad sense of the term.[16] The second architectural difference is the difference between domain-general and domain-specific architectures or, what we will take to be the same here (but see Samuels 1998; Cummins and Cummins 1999; Hirshfield and Gelman 1994), between a non-modular (or mildly modular) architecture and a massively modular architecture (Cosmides and Tooby 1987, 1997; Pinker 1997). We thus look at two opposing pairs of classes of architectures: classical vs. connectionist, and domain-general vs. domain-specific.[17]

6.1. *Connectionist vs. Classical Architectures*

As Fodor and Pylyshyn rightly see, the difference between classical and connectionist architectures is not about parallelism, speed, or biological realism but about *representational format*, in particular symbolic (or connectionist but local) representations vs. superposed and distributed representation, and about the cognitive processes afforded by each format. In what follows, we show (1) that differences in representational format imply normative differences relevant to assessing the rationality of cognitive processes, (2) that it is impossible to abstract away from differences in representational format and that, consequently, (3) systems that use solely symbolic representation and systems that use solely superposed and distributed representations belong to different epistemological strata.

We all share the intuition that while they are both members of the 'representation' family, instances of linguistic representations (languages, codes, etc.) and iconic representations (pictures, maps, etc.) belong to different representational formats, or *genera*, as Haugeland (1991) calls them.

It is usually thought that representational genera should be distinguished by the type of relation linking representations to their content (e.g., causality vs. isomorphism), but Haugeland argues that this confuses two distinct processes: recording and representation. The various genera should instead be distinguished by the structure of their content. A linguistic representation of an event represents (some of) the event's *absolute elements*, that is, elements that can be individuated independently of other represented element. An iconic representation of the same event represents (some of) the event's *relative elements*, that is, elements related to other represented elements in some space, real or abstract. According to Haugeland, distributed and superposed representation may belong to yet a third genus whose instances represent 'input-output associative' elements of the event, that is, *what should be done when*, or what Millikan (1995) calls "pushmi-pullyou representations".

The differences between the various representational formats are relevant to assessing the rationality of cognitive processes. Epistemic norms are at least in part about the management of representations: what representations a system should and should not hold (beliefs), when and what to add (learn, accept), delete (forget, reject), etc. Current models of rationality (good epistemic norms) all presuppose what we call *atomistic representation management*: the ability to pick out, evaluate, and add or delete *individual* representations. The models prescribe what rational systems *ought* to do with its representations, either individually or as groups of individuals (e.g., closed on inferential relations). And since *ought* implies *can*, the models presuppose that rational systems *can* manage their representations atomistically. But symbol users and vector users differ in their ability to manage their beliefs atomistically. Take the case of atomistic learning (adding a single representation to a knowledge base). The capacity is readily available, within limits, to systems that use symbolic representations, while it is utterly absent from systems that use fully superposed and distributed representations.

As long as there is enough memory (enough tape in the Turing machine), adding a symbolic representation simply means writing a new symbol (and deleting a symbolic representation simply means erasing or overwriting an old one). Because they represent absolute elements, symbolic representations have a digital character that makes the processes required for atomistic belief management transparent. Pushmi-pullyou representations do not. As a consequence, belief management in systems that use pushmi-pullyou representations is plagued by a form of catastrophic interference that makes them unable to sequentially add beliefs (McCloskey and Cohen 1989) and, in general, to manage their beliefs *atomistically*

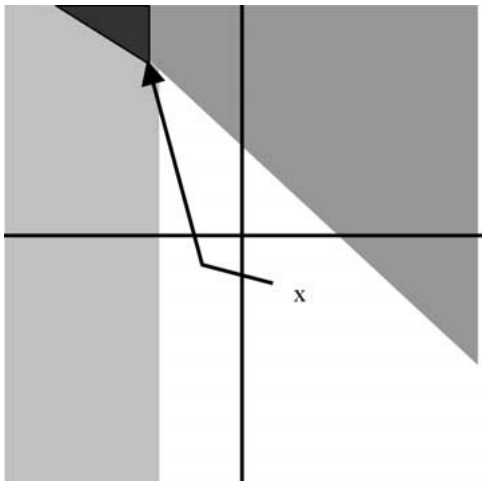

Figure 2. Concurrent learning.

(Poirier et al. 1999). Every connection in a net is involved in the production of outputs and new representations are added by adjusting the net's connections. But the very process of adding a new representation will alter all of the previously learned representations.

Learning in neural nets can be represented as movement in a 'solution space' (McCloskey and Cohen 1989), in which each point represents a possible configuration of weights. The set of configurations in which the net has successfully learned the required mapping (a pushmi-pullyou representation) is thus a region of solution space. Learning is thus any trajectory in solution space in which the net goes from a region where it does not appropriately map inputs and outputs to a region where it does. To train a net to represent two or more i/o mappings (i.e., two or more pushmi-pullyou representations), connectionists normally use a training set that contains them all; a training regime McCloskey and Cohen (1989) call 'concurrent learning' (Figure 2). It is well known that a number of algorithms for concurrent learning can nudge most nets to the appropriate region of solution space (where each mapping's region intersect). Take the case of two representations. Since they are presented with instances of both representations to be learned, concurrent learning algorithms possess all the relevant information to direct the trajectory towards the appropriate region of solution space. Movement in solution space cannot be blind to the location of the intersection and the only way to tell the algorithm which direction to take is to present instances of every mapping to be learned.

By contrast, atomistic learning involves learning the two representations sequentially, that is, learning the second mapping *without the use of*

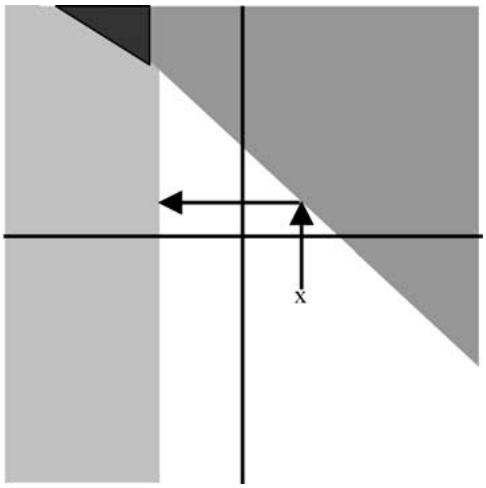

Figure 3. Atomistic learning.

the training data relevant to learning the first. In Poirier et al. (1999), we call this form of atomistic learning "conservative" and it is obvious that a lot of human learning, and forgetting, is conservative in this sense (see also Cohen and Eichenbaum 1993; McClelland et al. 1995). It should also be obvious from the solution space below (Figure 3) why conservative atomistic learning is impossible. Since it is only presented with one mapping at a time, the algorithm will take the short route from any current point in solution space towards the solution space for the new mapping, regardless of the general direction of the intersection and regardless of previously acquired knowledge. Any atomistic learning exhibited by the system will be purely accidental, for example, when the short route happens to push the net in the direction of the relevant intersection.

Atomistic learning is impossible for vector users (except by accident). Superposition builds an order and geometry in pushmi-pullyou representations that is anathema to atomistic belief management. As a consequence, any set of norms of rationality, any set of Rules of Right Reason, that presupposes atomistic belief management (that is, all current models of rationality) cannot apply to systems using pushmi-pullyou representations. These models presuppose an ability absent from these systems. But we took it as uncontroversial that, within limits, systems that use linguistic representations trivially have the capacity. Hence, the two kinds of systems differ in their ability to atomistically manage their beliefs. Since norms for vector users cannot require the impossible, their epistemology cannot be the same as the epistemology of symbol users (i.e., any system whose architecture is defined over symbolic representations).

But perhaps it is possible to abstract away from differences in representational format. Indeed, if the differences between representational genera are to be conceived in terms of the relation linking representations to their content, it is always possible to abstract away from representational differences: just make sure that both representations are about *the same thing*. If a picture and a descriptive sentence represent the same thing, then their representational differences disappear at the extensional level. Since the picture and the sentence are extensionally identical, it is always possible to abstract away from their intensional differences by "going extensional". But if, as Haugeland (1991) suggests, the representational differences between linguistic, iconic and pushmi-pullyou representations are conceived in terms of the different structural aspects of the contents they are sensitive to, then it is impossible to abstract away from representational differences by "going extensional". Linguistic representations cannot represent relative elements of a given content just as iconic representation cannot represent its absolute elements. "Translating" a representation from one genus to another *removes* the proprietary elements of the first and *adds* those the second, and thus one ends up with representations necessarily representing different *aspects* of the content.[18]

Connectionist and classical architectures use representations that belong to different genera and these differences, from which it is impossible to abstract away, imply normative differences relevant to assessing the rationality of the processes they realize. It follows that systems that have an exclusively classical architecture and systems that have an exclusively connectionist architecture belong to different epistemic groups. Thus, if brains exclusively use connectionist representations while individuals and everything above in the previous figure, including science, use linguistic representations, then brains and science belong to different strata.

6.2. *Domain-General vs. Domain-Specific Architectures*

It is more difficult to spell out exactly what distinguishes domain-general and domain-specific architectures because the nature of domain specificity is not well understood (Hirschfeld and Gelman 1994). To a first approximation, a domain (for a system) is a set of recurring problems (faced by the system) that share some distinguishing properties. This is a bit vague but sufficient in the current state of the art as some researchers individuate domains rather coarsely (vision, language, social relations, etc.), while others individuate them much more narrowly: face recognition, spatial relations, rigid object mechanics, tool-use, fear, social-exchange, emotion-perception, kin-oriented motivation, etc. (Tooby and Cosmides 1992). Consequently, an architecture is domain-general if it is designed to solve

problems from various domains. To do this, the architecture must be able to represent the domains' relevant elements and process information in ways that track its relevant structure. By contrast, an architecture is domain-specific if it can solve problems from only one domain, which means that it needs only represent elements from that domain and processes information to that track that domain's relevant structure. The difference between a domain-general and domain-specific architecture is the difference between, e.g., GPS[19] and an electronic chess player. GPS can solve problems from a variety of domains while the chess player's "cognitive system" only solves one problem: capturing the opponent's king.

This simple example is sufficient to show that domain-specificity *tout court* is a non-starter for human cognitive architecture. However coarsely domains are individuated, there are bound to be more than one and humans must be able to adequately deal with each. What is needed in the human case is *domain-specificity with breadth*. A system has breadth if it can solve problems from a variety of domains. There are two ways to build breadth in systems. Breadth is obviously an automatic property of domain-general architectures, so one way to build breadth is to simply build a system with a domain general architecture. Another way is to multiply domain-specific subsystems (e.g., modules), hence the current popularity of massively modular architectures. But in both cases, there is a price to pay for breadth. In the case of massively modular architectures, the price is coordination: how do you get all these modules to "talk" to each other and work together if each is optimized to work on one single kind of problem. One popular answer is that you don't always and the limitations of the human cognitive system exhibit just those failures. And, as Cosmides and Tooby (1987, 1994) and Tooby and Cosmides (1992) argue, *weakness*, that is the inability to deal with any but a domain's most simple problems, is the price of breadth in domain-general architecture.

A computationally complex problem is a problem that may give rise to a combinatorial explosion. Researchers in AI have known for a long time that combinatorial explosions cannot be prevented by adding more resources (space or time) since the problem's solution will often exceed any relevant space-time frame (one brain and 80 years in the case of humans[20]), and, in some famous cases (e.g., chess), will even exceed the resources of the known universe. The only way to prevent an explosion is to restrict the number of alternative hypotheses the system has to look at. And the only way to restrict the number of alternatives is build knowledge about the domain into the architecture, either in the form of 'innate' data structures or in the form of heuristics (heuristics make assumptions about the structure of the problems to be solved). Hence the popularity a while back of *micro-*

world in artificial intelligence. But whereas some philosophers have taken micro-worlds to be a bad thing (Dreyfus 1979), proponents of domain-specific massive modularity see it quite differently: micro-worlds are good provided each relevant (e.g. adaptively) micro-world is covered by its own private (micro-) domain-specific module.

In short, there is a tradeoff between robustness and breadth. Domain-general architectures have breadth but are often weak; domain-specific architectures are robust but have no breadth. As a result, the set of problems an architecture *can* solve varies as a function of it position on the 'general-specific' continuum. Since a set of norms cannot require a system to solve problems it cannot solve, norms that apply to systems with domain-specific architectures will necessarily be different from norms that apply to systems with domain-general architectures. The epistemology of domain-specific systems will necessarily be different from the epistemology of domain-general systems.

Finally, it is easy to see why it is impossible to abstract away from this architectural difference. If you take a domain-specific architecture and remove its domain specific knowledge and heuristics, you may turn it into a domain-general architecture[21] but the system will lose its original robustness. If you take a domain-general architecture and build domain specific knowledge, you will turn it into a robust domain specific architecture but lose its original breadth. It is impossible to abstract away from the difference while maintaining the relevant epistemological properties of the system, that is, in this case, the set of problems it *can* solve and, hence, the set of norms that will apply to the system.

7. WHY STRATA MATTER: BRAINS, INDIVIDUALS, AND SCIENCE

So far, we have argued for a point and moral:

- The point is that there is epistemic variability, especially in the vertical dimension. Like chocolate cake, epistemology is best when layered.
- The moral is that there are no Rules of Right Reason that apply across the board, and especially not across strata, and that forgetting this (or denying it, or simply being unaware of it) is a recipe for bad cognitive science and bad epistemology.

Although we have made both the point and the moral generally, we are especially interested in three strata: brains, individuals, and science. The fact that they belong to different strata means that the vocabulary, especially the normative vocabulary, used to describe problems and solutions

(i.e., mechanisms) at one level is likely to differ from the (normative) vocabulary used to describe problems and solutions at another. The price for not recognizing this fact is (possible) massive mischaracterization of systems residing at these levels. We say "likely" and "possible" because the sole fact that they do belong to different strata does not *entail* that the proprietary vocabulary of one level, or the mechanisms discovered at one level, are *necessarily* distinct from those of another, though it should by now strike you as wildly implausible that they would not be. It is still possible, however, that the solution space for a given type of epistemological problem is so small, or the landscape in that space so rough, with deep grooves and steep peaks, that all systems that do get to solve problems of that type will do so in essentially identical ways. But this kind of identity, if there is to be one, should be the result of research in cognitive science and philosophy of science and not something assumed beforehand in order to do that research. In this section, we want to emphasize the moral by showing how bad things can get (and have gotten) in philosophy and cognitive science when the moral is forgotten.

Take the case of the visual system, which is a nice example because it can be, and has been described at all three levels and thus provides a rich source of possible and actual confusion.[22] First let's stipulate a few things, some of which may be contentious. Our goal here is not to get things right as far as vision is concerned (nobody could at this point anyway since vision science is still evolving) but show how things could be: (1) Viewed from the level where science belongs, the function of the visual system is observation and evidential justification and its goal is truth (not accuracy since accuracy is graded). Visual knowledge at this level is justified true visual belief. Visual knowledge as justified true visual belief may be propositional (categorical, conceptual) and perhaps even sentential, since observation must be intersubjective. (2) Viewed from the individual level, the function of the visual system is belief fixation and conscious representation, in particular the construction of a 3-D conscious representation of the local mid-sized environment, and its goal is real-time processing, reliability and accuracy (hence the value of glasses and 20/20 vision when driving a car – by the way, that's why also you shouldn't drink and drive). Visual knowledge at this level is reliable and accurate on-time belief fixation and conscious representation. Visual knowledge as reliable and accurate on-time belief fixation and conscious representation is certainly not sentential and maybe not even propositional, though this issue is hotly debated, but rather iconic, metric, dense and analog. (3) Viewed from the brain level, things may be radically different, as there may not be such a thing as a visual system. Some cognitive scientists

(e.g., Ballard 1991; see also Churchland et al. 1994) have indeed argued that "the visual system" may not be a system at all, but a motley crew of various autonomous systems (modules), the vast majority of which are not involved in the construction of a conscious 3-D representation but in quick-and-dirty perception-action loops (Arbib 1981), or "intentional arcs" as foreseen by Merleau-Ponty (1945). Note that, on this view, these visual modules are not *components* of a visual *system* since they all work more or less independently, making the whole an aggregate (Wimsatt 1986) more than a system as such. If this view of vision at the brain level should turn out right, then each of these modules is a little visual system in its own right with its very own function. Visual knowledge at this level is successful motor control and representation construction (we stipulate). Knowledge as successful motor control is probably not iconic though it is probably spatial (dense, analog) in the sense in which weight matrices are. Visual knowledge as motor control may even be temporal (i.e., possess internal temporal structure) if the networks involved are recurrent (Churchland 1995).

Let's agree that epistemology is stratified as we suggest and that the visual system(s) is best described at different levels as we just did. To see the sort of confusion that may result if this is denied, let's focus on a few selected examples. In a nutshell, the coming story is about a series of mischaracterizations of the visual system and the problems they have caused in both philosophy and cognitive science.

7.1. *Confusing the Visual System as Described at the Science Level and the Same System as Described at the Individual Level*

The mistake here is describing belief fixation and conscious, visual representation of the environment as observation and evidential justification, and evaluating conscious representation with regards to truth and justification, or, conversely, describing observation and evidential justification as conscious representation, and evaluating their successes and failures with regards to individual psychological properties.

Consider observation. Philosophers of science like Hanson (1958) or Kuhn (1962) argued that observation is theory-laden and, hence, very plastic (change your theory and you change your observations). Fodor (1983, 1984) believes however that theory changes do not affect what is observed but only the beliefs we form on the basis of these observations. This disagreement, however, may simply be a verbal dispute: Fodor and the proponents of the theory-laden character of observation may both be right, their only mistake being that they think they're talking about the same thing – as you'd expect when different strata are not suitably distinguished.

Distinguishing strata properly, we see that Hanson and Kuhn understand observation as an science-level process. And we see that Fodor understands observation as a brain *and* individual level process: observations are the outputs of a modular (i.e., especially, informationally encapsulated and hardwired), sub-personal, brain-level process that are consciously available to individuals as appearances. Philosophers sensitive to the presence an epistemic layering of systems will not be tempted to see any substantive issue here. They will readily see that the debate isn't over observation but over "observation", that is, the proper use of the term: should we restrict the term to outputs of sub-personal modules and appearances, as Fodor suggests, or to the science-level process, as it is traditionally used in anti-empiricist philosophy of science. It might look like both constructs (Fodor's appearances and philosopher of science's observations) are being asked to play a role at the same epistemic strata. But it is far from evident that appearances (or qualia) play the distinguished role in scientific knowledge that observations were asked to play and against which Kuhn and Hanson mounted an attack (the fact that the sun appears to revolve around the earth, and that the latter appears no bigger than the moon, do not play significant roles in astronomy). Perhaps, rather, consciously available appearances belong to the individual level and that one of the great lessons learned by science is that appearances can never play a significant role in scientific knowledge. Fodor seems to realize this when he distinguishes observation from perceptual beliefs (Fodor 1984). But, once again, the debate then threatens to turn into mere verbal dispute. Maybe philosophers of science mean justified, true visual belief when they say observation whereas Fodor simply means appearances.

Or again, take the process of belief fixation. Many have argued that belief fixation is reliable but not necessarily good enough for justificatory purposes. Belief fixation is thus good, though not perfect, when evaluated with individual level norms, but much less so at the science level. Here's a way to get justified beliefs out of reliable but imperfect belief fixation mechanisms: Enforce a "not seen by anyone unless seen by everyone in relevantly similar conditions" policy.[23] When this policy is enforced, as it has been for a few centuries now, apparitions of the Virgin Mary do not get to justify the belief that she exists and bubbles along electrodes do not get to justify the belief that there can be nuclear fusion at room temperature. Parapsychology remains without evidential basis for similar reasons. But by definition, more than one individual is needed to get observational evidence for beliefs through strict enforcement of this policy and thus this kind of evidential justification is not an individual epistemic capacity.

The converse mistake, confusing observation and evidential justification with conscious representation is a hallmark of positivistic philosophy of science and much of empiricism. Sense data and protokolsätze both have this dual nature: they are supposed to both be experienced by individuals (or sentences individuals could accept on the basis of their experiences in the case of protokolsätze) and justify scientific knowledge. As the repeated failures of positivism and empiricist philosophy of science suggest, it seems reasonable to think that no single event that could have both properties. So is empiricism a story about the origin and justification of individual knowledge or scientific knowledge? On our view, there *could* be an empiricist story to tell about knowledge at the individual level, and there *could* an empiricist story to tell about knowledge at the level where science belongs, *but these won't be the same* (and of course one may be true and the other not). Indeed, if we follow Fodor and Pylyshyn (1988) in thinking that connectionism is just plain old associationistic empiricism in modern, computer-simulated, clothes, then empiricism is also a distinct possibility at the brain level. But, once again, this does not mean that it is the same kind of empiricism and that the failures of empiricism at the individual or science level should be taken as inductive evidence for the eventual failure of empiricism at the brain level. It may yet turn out that Hume basically got it right as far as the brain goes (though his notion of association is shallow by today's standard) but got into deep trouble by not properly distinguishing the three strata.

7.2. *Confusing the Visual System as Described at the Individual Level and the Same System as Described at the Brain Level*

The mistake here is describing the neural processes involved in vision as mechanisms for belief fixation and for the construction of a conscious representation, and evaluating their successes and failures of these processes with respect to conscious visual properties and folk psychological norms and properties. Or conversely the mistake is describing belief fixation and conscious representation as neural processes.

Traditional work on vision in both cognitive science and AI may have suffered from that confusion: the computational description of the task (e.g., in Marr's theory) is lifted from folk psychology and quotidian (as opposed to trained) phenomenology. To emphasize the point, Ballard (1991, 57) quotes from the *Encyclopedia of Artificial Intelligence*: "The goal of an image understanding system is two transform two dimensional data into a description of the three dimensional world" and such a system "must infer 3-D surfaces, volumes, boundaries, shadows, occlusion, depth, color, motion". In short, the function of neural processes involved in vision is

to transform patterns of retinal activation into the 3-D Technicolor movie we experience. But according to Ballard, and Churchland et al. (1994), this may wholly mischaracterize the function of vision at the neural or machine level. It is widely recognized now that quotidian phenomenology, especially visual phenomenology, is an illusion, one that is easily disturbed by injuries or drugs. The fact that we do experience this illusion makes it an *explanandum* for neurology but this does not entail that neurology should lift its description of the function of vision from quotidian phenomenology. Ballard's animate vision paradigm, for instance, centers on gaze control and he argues that the existence of such mechanisms fundamentally changes the computational task of visual systems. But gaze control is all but absent from quotidian phenomenology; certainly it isn't seen at that level as a fundamental determinant of the computational task of visual systems (rather only as something we do to d'viewd'd' something else). It is not important here whether these researchers are right or not to describe vision as they do. The simple fact that the need is felt to distinguish the brain (or machine) level description of vision from its individual level description (as given by phenomenology and folk psychology) is sufficient to prove our point: importing vocabulary from one stratum to describe systems at another may generate more confusion than understanding.

The point cuts both ways, however. Just as it may generate confusion to import individual level descriptions at the brain level, it may be an error to import brain level descriptions at the individual (or science) level. Some eliminativist arguments may commit such an error. It does not follow from the fact that individual-level, folk psychological visual properties (e.g., conscious representation and belief fixation) do not reduce to neurological properties of the brain, that they should be eliminated. Conscious visual properties (visual *qualia*), for instance, may depend on the brain but also on a stable field of vision (proximal environment) and the ability to saccade (eye movements), which requires non-neurological body events to occur (muscles flexing). Immobilizing the eyes or getting the environment to move in step with the eyes both disrupt conscious visual perception. Situatedness and embodiment may be necessary for individual level visual properties to occur. Note that, on this view, the relationship between visual *qualia* and muscle movements and the visual field is not the same as that between visual *qualia* and, say, oxygen.[24] The presence of oxygen is, of course, necessary for visual experience (absence of oxygen causing death), but, according to the embodied cognition view, the external field and motor movements are part of the cognitive mechanism that generates visual experience.[25]

8. NEW PROBLEMS AND PROSPECTS

Our argument up to this point is that epistemology, if it is to have a subject matter, must be pluralistic. Pluralism takes form in the many epistemic strata discussed in Section 2. Stratified epistemology raises novel questions and problems that must be addressed if it is to become a useful tool for theoretical and practical analysis of epistemic systems.

The first cluster of issues and questions raised by stratified epistemology concerns the nature and status of these strata and how they are related. For example, what are the necessary and sufficient conditions for genuine "levelhood"? We have talked about levels being constituted by architectures, resources, and goals, and these certainly will become a part of any complete analysis of levels. A more difficult issue is how fine-grained do we want our analysis of levels to be. We have mentioned that science and brains are an obvious example of two different strata, since they differ in architecture, resources, and goals. But are two different brains of the same species, e.g., two different human brains, members of the same strata? Architecturally they are bound to be quite similar, and the same goes for resources and goals. However, there will be differences (in some cases, quite vast differences). At what point do we exit one stratum and enter another? Similarly, does the institution of science as it was in 17th century England belong to the same stratum as 21st century science as it is practiced in the US? It is obvious that the resources (and probably architecture) of the two periods are quite different.

One way that might help us distinguish strata is to ask whether certain epistemic concepts apply to the system(s) under consideration. For example, beliefs, whose contents may be something like propositions, can be true, false, justified, etc. But truth may not be the appropriate concept when applied to the deliverances of the visual system; a graded notion like accuracy seems better suited. Also, it is not clear in what sense a picture is justified. For that matter, it is unlikely that the concepts of justification or evidence, as traditionally conceived, will apply to the elements that make up weight matrices and activation spaces in any natural or straightforward way. This may be a clue that a system whose architecture consists of weight matrices and the like belongs to a different strata than a system whose architecture consists of propositional attitudes.

Another issue concerns the relationship between the constituent hierarchy and its organization and the various strata and their organization, and how cognitive systems at higher levels may be realized by cognitive systems at other levels. For example, we believe science is composed of, among other things, individuals, and individuals are composed of brains.

This is a part/whole ordering of architectures. We assume, furthermore, that the component architecture(s) will constrain the way the higher levels that they compose will operate. In this way, there is a natural dependency, *epistemically*, that mirrors the compositional dependency. However, there appear to be top-down dependencies, too. Part of my resources as an individual includes cultural artifacts like instruments, tools, books, etc. Language, too, may be part of the architecture of an individual and also something that requires a larger context, cultural or otherwise, to acquire. In this way, the resources and architecture of an individual may depend on a "higher" stratum. Another example of the way "higher" strata can affect "lower" strata comes from evolutionary considerations. If adaptation is a sort of long-term learning, and describes the way a species learns, then the architecture, resources, and goals of future individuals will depend on this higher order epistemic process.

The above examples raise the general issue of understanding how the various strata interact and constrain each other. Specifically, if the various strata each have their own goals, and if the goals are in competition or conflict, how do the conflicts get resolved? In other words, is there a general model for conflict resolution that can apply to all the levels? Along a similar vein, it is obvious that there has to be some sort of coordination between the various strata, such that the activities at each level work in lock-step. For example, in order for science to get done, in order for the goals of science to get accomplished, certain demands placed on individuals must be satisfied, e.g., time commitments involving physical and mental exertion, cooperation with fellow scientists, etc. This seems to require putting other demands from other strata on hold. Failure to do so can have heavy consequences (e.g., sexual harassment at the workplace). What is needed, then, is a model for how activities get coordinated, how goals get subordinated and prioritized, and so forth.

A second cluster of issues concerns what the goals of the various strata are, and where they come from. Our normative evaluation of a cognitive system depends on assessing its performance in relation to what it is supposed to do, or what its goals are. In considering the goals of a cognitive system, we should keep the following in mind:

1. There may be a distinction between the goals of a system and its function. For example, we might suppose that the visual system functions to help us avoid being killed by predators, and a visual system is "good" or is performing well insofar as it achieves this function. In order to achieve this function, however, the visual system may be designed to construct highly accurate spatial representations of the local environment. If this is the case, we may say the goal of that particular

system is to provide accurate representations of the local environment. Notice that the system may be quite effective at achieving its goals, but less good at achieving its function (it's too slow for example), or vice versa. A similar point can be made about science. Perhaps science functions to give its practitioners and consumers some selective advantage or other. As a means of achieving this function, perhaps science has as its goal truth or understanding. In such a case, science may succeed at the latter while fail at the former. Prima facie, we have a reason to keep the two separate.

2. A system may have multiple goals, sub-goals, etc. There is no reason to suppose that there is only one goal at any strata.
3. The goals/functions (g/f) at one stratum may be fixed by the g/f at other strata. This is another way in which the strata affect one another. The g/f of the visual system may be fixed by the g/f of the individual, for example (the goals of the visual system may be fixed by its function, where the function is fixed by the g/f of the individual). Also, there may be one strata whose g/f set the g/f for all the other strata.

Stratified epistemology also may help us conceive differently the traditional project in philosophical epistemology. We all learn that epistemology is the theory of knowledge and that knowledge is justified true belief (plus or minus a bit). But another picture of epistemology emerges if we insist that norms are constrained by capacities and that some of these constraints are ineliminable (and make sure that the relevant concepts – justification, truth, belief – are used unambiguously from system to system). Some systems do not aim at justified beliefs but simply do not want to be misinformed (see e.g., Goldman 1999). Other systems are not primarily concerned by truth, but speed or real-time action matters a great deal to them (Clark 1996). They want information as accurate as possible, but will settle for less that truth given their time constraints. Some systems *cannot* change their beliefs about the environment in the face of incoming evidence (they literally cannot change their mind). But over the course of many generations, new evidence will impact on the organism's species in such a way that their descendents will harbor more accurate beliefs about the environment. Other systems that do act based on information they collect from the environment cannot have beliefs at all, if beliefs are anything like discrete, truth-valuable, pieces of information. Yet most would say that these systems deal, in some sense, in the commodity we call knowledge. If this is true, there is more to epistemology than the theory of justified true beliefs. The reductive view that identifies epistemology with the theory of justified true beliefs is part and parcel of the problem we denounce here: the collapse of the vertical dimension. Perhaps justified beliefs are

the relevant construct to understand knowledge at the institutional level (e.g., science). And perhaps veritistic epistemology is relevant everywhere individuals are concerned. But this does not mean that these constructs apply to other strata. Justification may be normative in science whereas individuals may simply wish to avoid misinformation (to them, justification may just be a pragmatic means of minimizing misinformation).

The proponent of a more traditional view will rightfully ask: what is epistemology about then? This is a difficult question in light of the vertical worry we address but, to a first approximation, we may say that epistemology is a theory of normatively constrained agency (where agency is both behavior and the informational structures that guide it). Ethics is also a theory of normatively constrained agency. That is why epistemology and ethics have so much in common: why so many epistemological problems and solutions have a counterpart in ethics. The one thing that does distinguish them, however, is the type of norms they appeal to: norms for rational agency vs. norms for moral agency. Accordingly, to a first approximation, we may then say that epistemology is the theory of rationally constrained behavior and information bearing. Traditional epistemology focuses on one informational structure (belief) and two sets of norms (truth and justification) and that is OK. But we claim there is more to epistemology.

Another consequence of stratifying epistemology may be to expose the theoretical/practical reason distinction as being too crude. Our position is that the effectiveness of a system trying to achieve some goal (its acting) is the thing to be assessed, but that there is a reasoning component to this that can be evaluated for accuracy; in other words, we believe that there is a connection between the theoretical reasoning and practical action.[26] However, the problems and issues surrounding the relationship between the theoretical and the practical are framed too syllogistically and discursively, focusing on truth and justification at the expense of other ways of conceiving of the connection between reasoning and action.

Finally, once we take stratified epistemology seriously, the alleged incompatibility of folk psychology with neuroscience might dissolve if we recognize that the folk have different resources, goals and architecture than their brains; at minimum, folk psychology takes the unit of information to be belief, where such beliefs can be atomistically managed, that is, added and deleted individually, and can occupy roles in inferential structures of the type described by first order logic. Brains do *not* appear to work this way. Connectionist cognition appears to be radically different from the models proposed by both GOFAI (Good Old Fashioned Artificial Intelligence; Haugeland 1985) and folk psychology. The incompatibility, and

thus the problem, arises when folk psychology is taken as the model of cognition *generally*, and thus, the model for brains specifically. If folk psychology and neuroscience are explanatory strategies that apply to different strata (they describe different forms of reasoning at different levels), then the incompatibility may be more apparent than real.[27]

9. CONCLUSION

The development of cognitive science has broadened our conception of rationality beyond conscious, human adults, and has thereby expanded the range of things subject to normative epistemic assessment. Until recently, however, the theory of rationality has been couched in the language of propositional attitudes. Connectionist models have made us aware that cognition is not essentially tied to the propositional attitudes, and that cognitive systems can be quite diverse: they can differ with respect to architectures, resources, and goals.

This cognitive diversity has important consequences for the normative assessment of cognitive systems. Since ought implies can, we cannot require a system to do something for which it is incapable. Systems with diverse architectures, resources and goals will have different capacities, and thus the norms they are subject to will also be diverse. This diversity is embodied in stratified epistemology.

Failure to recognize this cognitive diversity, and the diversity of Rules of Right Reason that come with it, has led to many possible confusions in cognitive science and philosophy of science. Our conclusions regarding this matter are:

1. Scientific cognition is not a good model of cognition generally (contra early Fodor, early-Chomsky, Gopnik, etc.), because cognition is essentially diverse.
2. Therefore, the Rules of Right Reason that apply to science (scientific rationality) do not apply to individuals or brains, since science, individuals, and brains are different kinds of cognitive systems, and the Rules are constrained by the kind of cognitive system under consideration.
3. The same applies to folk psychology and connectionism. They are not good models of cognition generally, either, for the same reason given in 1; therefore, the Rules of Right Reason that apply to them will not apply across cognitive systems, for the reason given in 2.

The trouble occurs when we attempt to identify the forms of reasoning used in science with the forms of reasoning used by individuals and brains.

We are now in a position to see why this is a mistake. Science, with its attendant resources and goals, constitutes only one of a variety of cognitive systems; it occupies one of many epistemic strata. Since the resources and goals of science are different from the resources and goals of other strata, the forms of reasoning used in science and the norms which govern scientific reasoning will typically *not* apply to the other strata. As a result, those theories that collapse the distinctions among the strata are apt to mischaracterize the resources, goals, and norms at the various strata, resulting in both bad philosophy of science and cognitive science.

NOTES

[1] See Cummins (1995) for an argument that the rationales required for understanding cognitive processes in neural networks can not be understood in traditional terms.

[2] See Cummins et al. (2001), for a discussion of the difference between encodings generally and structural representations.

[3] Talk of 'dimensions' is simply an aid to understanding what we mean by epistemological strata and should not be taken too literally. While the vertical dimension is indeed a dimension (at least an ordering, though not a metric), the horizontal dimension is really just a set, or cluster, of systems, and not an ordering of anything. In fact, we are quite open to the possibility that some of these sets may contain only one member, though we believe that some important strata contain more than one. To give a few examples, we believe that the neurological stratum may contain a number of systems with different representational formats, computational architectures and resource constraints. If massive modularists are right, each domain specific module might count as a distinct member of the cognitive stratum . Finally, the formal group or institutional level may contain groups with varying resources constraints: the judicial system, part of which may understood as an epistemic system whose function it is to determine who committed a crime, has more resources than a PI team trying to achieve the same goal.

[4] Fodor still believes that psychology is "philosophy of science writ small" (Fodor 2000, 52) and the reason he gives (it's wildly implausible that human cognition has changed in the last few hundred years) perfectly illustrates our point: If you recognize the vertical dimension, human cognition does not have to change to give rise to science. It only needs to be co-opted into a larger system: instead of using your visual acuity to spot ripe berries, use it to spot traces of bosons in a bubble chamber and then use some of the money you receive for doing that to buy ripe berries. See below for the various functions of the visual system in individual and scientific cognition.

[5] It is important to distinguish, in this context, the cognition of individual scientists from scientific cognition. Is hypothesis formation and confirmation the kind of thing that occurs in an individual scientist's head or something that occurs at the level of science as an institution? I can form the hypothesis that I have $0.54 in my pocket and confirm it by looking. But cases of that sort are rare in science as it is practiced today – where, to paraphrase Hillary Clinton, it takes a village to confirm a hypothesis (take any hypothesis in physics for instance, or the workings of any modern lab. See Galison (1987) for a detailed example of this)). At the level of science, there are standards or norms that are not established by the person who formulates the hypothesis but pre-determined, as it were, by others (hence the

village). If those rules say that an hypothesis has not been confirmed unless the confirming evidence has been replicated by some other lab or scientist (a village again), then clearly the process of confirmation does not occur at the individual level. Consider the hypothesis of Universal Grammar (UG) and assume it is true. Was the hypothesis confirmed by Chomsky himself or by the work of a generation of linguists of Chomskyian persuasion?

[6] Hutchins (1995a) argues that modern Western navigation can be viewed as a cognitive system in its own right, above and beyond the individual cognitive agents that compose it. If the practice of navigation can be successfully modeled as a cognitive system, it is not a stretch to suppose that science as an institution can be so modeled as well.

[7] It is natural to wonder how systems at the various strata are realized – e.g., whether they or their components are realized as systems from lower strata. We leave this question aside in order to concentrate on the positive reasons for accepting what we are calling vertical epistemological pluralism.

[8] The effects of insufficient or inaccurate information might be taken into account in the theory of ideal rationality itself.

[9] While there are various ways in which cognitive systems can expand (or contract) their memory, or at least allocate more (or less) space to a given task, more memory generally requires more computation time. It takes longer to search (and otherwise manage) a larger space. Greater density can compensate for this effect only within limits before accuracy is compromised.

[10] We owe this last example to Paul Teller.

[11] Distinguishing the important cases also seems to require reference to a containing system: what is it important to see? Which scientific problems is it most important to solve? The question arises: important for what? The progress of science may require solutions to fundamental problems, whereas the welfare of society may require solutions to pressing practical problems. (The time frame will be important here.)

[12] There are trade-offs when more than one consumer is involved. High blood pressure gets oxygen to cells faster, but increases the risk of stroke.

[13] This seems to be what is missing in the classic social psychological literature (Kahneman and Tversky 1973; Nisbett and Ross 1980), which measures rationality in terms of strategies for expected success.

[14] That humans are subject to the Gambler's Fallacy is often cited as evidence that humans are irrational. But in a world where most events are not independent, it is a pretty effective strategy. It may seem more rational to determine whether we are in a situation where events are independent, and then adopt the strategy that is most appropriate for that situation. But devoting the computational resources to determine this will require taking computational resources away from other tasks, and doing so may do more harm than good.

[15] The frog's bug detector may be able to 'think' "I want to eat this" but not "this wants to eat me" while its "large-mammal (or Frenchman) detector" may be able to 'think' "This wants to eat me" but not "I want to eat this".

[16] If he did, the upshot of his pessimism about theories of central processing (Fodor 1983, 2000) would be that there is no, and may never be any, cognitive science *tout court*. This is surely too strong and a more reasonable claim would be that there is no, and may never be any cognitive science *of thought* (see Cummins 1989 for a similar worry).

[17] In what follows, we only discuss brain- and individual-level examples, since these are the most widely researched as cognitive systems; but there are other, albeit more speculative, examples. Discussing 'how a cockpit remembers its speed' Hutchins (1995b) clearly sees the whole cockpit as a system possessing its own cognitive architecture: "This system

[the cockpit system] makes use of representations in many different media. The media themselves have very different properties. The speed card booklet is a relatively permanent representation. The spoken representation is ephemeral and endures only in its production. The memory is stored ultimately for use in the physical state of the speed bugs. It is represented temporarily in the spoken interchanges, and represented with unknown persistence in the memories of the individual pilots. The pilot's memories clearly are involved, but they operate in an environment where there is a great deal of support for recreating the memory" (p. 285). Similarly, many organization anthropologists think of the structure that generates and sustain information flow in companies and various institutions as a cognitive architecture whose components are libraries, memos, manuals, officially established (or sometimes implicit) procedures, and employees are components of the architecture (See Nonakata and Tageushi 1995). Finally, much discussion in the philosophy of science could be *construed* as attempts to unearth (or create, or justify) the cognitive architecture of the institutional-level cognitive process that is science. It should be noted that pilots in a cockpit, organizations and science all have different goals, time-pressures, memory constraints, etc., so that what we say here about architectures could be extended to these other determinants of a system's proper epistemic stratum.

[18] Note that this kind of "translation" is different from normal translation, say French to English, in which one starts and ends with representations of the same genus (linguistic) that thus represents the same aspects of the event (absolute). If one translates "Le roi est mort" into "The king is dead", one is talking about the same event, the death of the king, *and one is capturing the same structural aspects of that* event. But when one "translates" a picture into words, or vice versa, one is still talking about the same event, *but one is capturing different structural aspects of that event*: absolute vs. relative. If one "translates", say, a picture of a dead king into words, one may say "The king is dead". But one may also say "The tyrant finally got his come-uppance", or "Henry V succumbed to his wounds" or whatever, indefinitely. This looseness between the picture and its many "translations" is due to the fact that pictures do not represent absolute elements. The "translator" may choose any of a number of absolute elements to "translate" the picture's relative content. But the choice is a pragmatic affair that reflects the knowledge and interests of the "translator" and not a semantic affair entirely driven by the picture's content. In such "translations", the original (i.e., relative) content of the pictorial representation is lost and the new content added, the absolute content of the linguistic representation, will reflect the interests and knowledge of the translator. That the relative content of the picture is lost is easily seen by the fact that another "translator", unaware of the original picture and asked to translate the words back into picture, may create a picture whose relative content has absolutely nothing in common with those of the original picture (one picture may represent a royal figure lying on a battlefield and the other a ghost talking to a prince). See Haugeland (1991) for a more detailed argument.

[19] That is, Newell et al.'s (1958) General Problem Solver and not, of course, the Department of Defense's Global Positioning System.

[20] Actually, the most complex problems humans face must be solved in less than 5 years, that is, during early childhood: language, social relations, behavior prediction, etc.

[21] Most probably, you will just turn it into a piece of junk.

[22] We must be careful with ontology here so as not to generate some confusion or our own. We said that constituency relations order the various strata in a vertical fashion (that's why we called them strata, and not, say, "regions"). Since constituency is transitive, strata membership is, as were, "upwardly heritable": a system that belongs to a given strata will

also belong all higher strata, as is the case, here, with the visual system. But this does not entail that *functionality* is upwardly heritable. Systems will be components of *different* higher-level systems: the visual system is a component of an organism at the individual level and of an institution at the science level. And they will of course play different roles in these different systems, which, by our lights (Cummins 1975, 1983), means that they have different functions at these respective levels. The function of the visual system at one level may thus be different from, and (because of conflicting demands) perhaps even in tension with its function at a higher-level.

[23] "Relevantly similar conditions" are, of course, difficult to specify in a way that is not question-begging, because what counts as a relevantly similar condition is, to some extent, an empirical question which we tend to settle either by appeal to theory or by appeal to evidence that variability in a condition accounts for a significant portion of the variability in what is "seen".

[24] We thank one anonymous reviewer for pointing out this possible interpretation of our argument.

[25] In a recent BBS article, Kevin O'Regan and Alva Noë (2001) defend this view: "We propose that seeing is a way of acting. It is a particular way of exploring the environment. Activity in internal representations does not generate the experience of seeing. The outside world serves as its own, external, representation. The experience of seeing occurs when the organism masters what we call the governing laws of sensorimotor contingency" (p. 939).

[26] Traditional epistemology is right to point out that there is an accuracy component to reasoning, but wrong in supposing that it is the thing to be assessed when evaluating the rationality of the system. On the other extreme, those who stress effectiveness at the expensive of accuracy are right that effectiveness is the thing to be assessed, but wrong that there is no accuracy component. We are sympathetic to John Pollock's (Hume's originally?) view that the practical is prior to the theoretical, and that the theoretical is in service to the practical.

[27] A problem, of course, is that most consider folk psychology and neuroscience to be offering competing explanations for the same phenomena (memory, dreams, etc.), and thus they do look incompatible. This may be an artifact of one explanatory strategy trying to explain everything cognitive – the very thing we warn against. So perhaps both theories are trying to explain the same phenomena but ought not to try.

REFERENCES

Arbib, M. A.: 1981, 'Perceptual Structures and Distributed Motor Control', in V. B. Brooks (ed.), *Handbook of Physiology: The Nervous System II. Motor Control*, American Physiological Society, Bethesda, Maryland, pp. 1449–1480.

Ballard, D. H.: 1991, 'Animate Vision', *Artificial Intelligence* **48**, 57–86.

Calvin, W. H. and G. A. Ojemann: 1994, *Conversations with Neil's Brain*, Perseus Books, Reading, Massachusetts.

Chomsky, N.: 1965, *Aspect of the Theory of Syntax*, MIT Press, Cambridge, Massachusetts.

Churchland, P. M.: 1979, *Scientific Realism and the Plasticity of Mind*, Cambridge University Press, Cambridge.

Churchland, P. M.: 1981, 'Eliminative Materialism and the Propositional Attitudes', *The Journal of Philosophy* **78**, 67–90.

Churchland, P. M.: 1989, *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press, Cambridge, Massachusetts.
Churchland, P. M.: 1995, *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*, MIT Press, Cambridge, Massachusetts.
Churchland, P. S.: 1986, *Neurophilosophy: Toward A Unified Science of the Mind-Brain*, MIT Press, Cambridge, Massachusetts.
Churchland, P. S., V.S. Ramachandran and T. J. Sejnowski: 1994, 'A Critique of Pure Vision', in C. Koch and J. Davis (eds.), *Large-Scale Neuronal Theories of the Brain*, MIT Press, Cambridge, Massachusetts.
Clark, A.: 1995, 'Moving Minds: Situating Content in the Service of Real-Time Success', *Philosophical Perspectives* **9**, 89–104.
Cohen, N. J. and H. Eichenbaum: 1997, *Memory, Amnesia, and the Hippocampal System*, MIT Press, Cambridge, Massachusetts.
Cosmides, L. and J. Tooby: 1987, 'From Evolution to Behavior: Evolutionary Psychology as the Missing Link', in J. Dupré (ed.), *The Latest on the Best*, MIT Press, Cambridge, Massachusetts.
Cosmides, L. and J. Tooby: 1994, 'Origins of Domain Specificity: The Evolution of Functional Organization', in L. A. Hirschfield and S. A. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge University Press, Cambridge.
Cosmides, L. and J. Tooby: 1997, 'The Modular Nature of Human Intelligence', in A. B. Scheibel and J. W. Schoff (eds.), *The Origin and Evolution of Intelligence*, Jones and Bartlett Publishers, Boston.
Cummins, D. D.: 1995, *The Other Side of Psychology*, St. Martin's Press, New York.
Cummins, D. D. and R. Cummins: 1999, 'Biological Preparedness and Evolutionary Explanation', *Cognition* **73**, B37–B53.
Cummins, R.: 1975, 'Functional Analysis', *The Journal of Philosophy* **72**, 741–765.
Cummins, R.: 1983, *The Nature of Psychological Explanation*, MIT Press, Cambridge, Massachusetts.
Cummins, R.: 1989, *Meaning and Mental Representation*, MIT Press, Cambridge, Massachusetts.
Cummins, R.: 1995, 'Connectionism and the Rationale Constraint on Cognitive Explanations', *Philosophical Perspectives* **9**, 105–125.
Cummins, R.: 1996, *Representations, Targets, and Attitudes*, MIT Press, Cambridge, Massachusetts.
Cummins, R., J. Blackmon, D. Byrd, P. Poirier, M. Roth and G. Schwarz: 2001, 'Systematicity and the Cognition of Structured Domains', *The Journal of Philosophy* **98**, 167–185.
Cummins, R. and D. D. Cummins: 2000, 'Introduction to Part 1', in R. Cummins and D. D. Cummins (eds.), *Minds, Brains, and Computers: The Foundations of Cognitive Science, an Anthology*, Blackwell, Oxford.
Dempser, F.: 1981, 'Memory Span: Sources of Individual and Developmental Differences', *Psychological Bulletin* **89**, 63–100.
Dennett, D. C.: 1969, *Content and Consciousness*, Routledge and Kegan Paul, London.
Dennett, D. C.: 1978, *Brainstorms: Philosophical Essays on Mind and Psychology*, Bradford Books, Montgomery, VT.
Dreyfus, H.: 1979, *What Computers Still Can't Do*, Harper and Row, New York.
Elman, J. L.: 1993, 'Learning and Development in Neural Networks: The Importance of Starting Small', *Cognition* **48**, 71–99.

Elman, J. L., E. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi and K. Plunkett: 1996, *Rethinking Innateness*, MIT Press, Cambridge, Massachusetts.

Fodor, J. A.: 1968, 'The Appeal to Tacit Knowledge in Psychological Explanation', *The Journal of Philosophy* **65**, 627–40.

Fodor, J. A.: 1975, *The Language of Thought*, Crowell, New York.

Fodor, J. A.: 1983, *The Modularity of Mind: An Essay in Faculty Psychology*, MIT Press, Cambridge, Massachusetts.

Fodor, J. A.: 2000, *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, MIT Press, Cambridge, Massachusetts.

Fodor, J. A. and Z. W. Pylyshyn: 1988, 'Connectionism and Cognitive Architecture', *Cognition* **28**, 3–71.

Galaburda, A. M. and M. Livingstone: 1993, 'Evidence of a Magnocellular Defect in Neurodevelopmental Dyslexia', *Annals of the New York Academy of Sciences* **682**, 70–82.

Galison, P.: 1987, *How Experiments End*, University of Chicago Press, Chicago.

Giere, R.: 1988, *Explaining Science: A Cognitive Approach*, University of Chicago Press, Chicago.

Goldman, A. I.: 1999, *Knowledge in a Social World*, Oxford University Press, Oxford.

Goldowsky, B. and E. L. Newport: 1990, 'The Less is More Hypothesis: Modeling the Effect of Processing Constraints on Language Learnability', unpublished manuscript, University of Rochester, Rochester, New York.

Gopnik, A. and A. N. Meltzoff: 1997, *Words, Thoughts, and Theories*, MIT Press, Cambridge, Massachusetts.

Hanson, N. R.: 1958, *Patterns of Discovery*, Cambridge University Press, Cambridge.

Haugeland, J.: 1985, *Artificial Intelligence: The Very Idea*, MIT Press/Bradford Books, Cambridge, Massachusetts.

Haugeland, J.: 1991, 'Representational Genera', in W. Ramsey, S. Stich and D. Rumelhart (eds.), *Philosophy and Connectionist Theory*, Lawrence Erlbaum, Hillsdale, New Jersey.

Helmholtz, H. von: 1866, *Treatise on Physiological Optics*, 3rd edn., in J. P. C. Southall (trans.). Opt. Soc. Amer. New York, 1924. Dover reprint 1962.

Hirshfield, L. A. and S. A. Gelman (eds.): 1994, *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge University Press, Cambridge.

Hume, D.: 1739, *A Treatise of Human Nature*, 2nd edn., in L. A. Selby-Bigge and P. H. Nidditch (eds.), Oxford University Press, Oxford 1978.

Hutchins, E.: 1995a, *Cognition in the Wild*, MIT Press/Bradford Books, Cambridge, Massachusetts.

Hutchins, E.: 1995b, 'How a Cockpit Remembers its Speed', *Cognitive Science* **19**, 265–288.

Kahneman, D. and A. Tversky: 1973, 'On the Psychology of Prediction', *Psychological Review* **80**, 237–251.

Kitcher, P. S.: 1996, 'From Neurophilosophy to Neurocomputation: Searching the Cognitive Forest', in R. N. McCauley (ed.), *The Churchlands and their Critics*, Basil Blackwell, Oxford.

Kuhn, T.: 1962, *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

McClelland, J. L., B. L. McNaughton and R. C. O'Reilly: 1995, 'Why there are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures pf Connectionist Models of Learning and Memory', *Psychological Review* **102**, 419–457.

McClelland, J. L. and D. E. Rumelhart: 1988, *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*, MIT Press/Bradford Books, Cambridge, Massachusetts.

McCloskey, M. and N. J. Cohen: 1989, 'Catastrophic Interference in Connectionist Networks: The Sequential Leaning Problem', *The Psychology of Learning and Motivation* **24**, 109–165.

Merleau-Ponty, M.: 1945, *Phénoménologie de la perception*, Gallimard, Paris.

Millikan, R. G.: 1995, 'Pushmi-Pullyou Representations', *Philosophical Perspectives* **9**, 185–200.

Newell, A., H. A. Simon and J. C. Shaw: 1958, 'Elements of a Theory of Human Solving', *Psychological Review* **65**, 151–166.

Newport, E. L.: 1990, 'Maturational Constraints on Language Learning', *Cognitive Science* **14**, 11–28.

Nisbett, R. and L. Ross: 1980, *Human Inference*, Prentice-Hall, Englewood Cliffs, New Jersey.

Nonaka, I. and H. Takeuchi: 1995, *The Knowledge-Creating Company*, Oxford University Press, London.

O'Regan, K. and A. Noë: 2001, 'A Sensorimotor Account of Vision and Visual Consciousness', *Behavioral and Brain Sciences* **24**, 939–173.

Pinker, S.: 1997, *How the Mind Works*, Norton, New York.

Poirier, P., R. Cummins, J. Blackmon, D. Byrd, M. Roth and G. Schwarz: 1999, 'The Epistemology of Non-Symbolic Cognition: Atomistic Learning and Forgetting', Tech. Report Phil99-3, University of California-Davis.

Pollock, J.: 1989, *How to Build a Person: A Prolegomenon*, MIT Press, Cambridge, Massachusetts.

Putnam, Hilary: 1962, 'The Analytic and the Synthetic', in H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science*, Vol. III, University of Minnesota Press, Minneapolis, pp. 350–397.

Samuels, R.: 1998, 'Evolutionary Psychology and the Massive Modularity Hypothesis', *British Journal of Philosophy of Science* **49**, 575–602.

Tallal, P., S. L. Miller, W. M. Jenkins and M. M. Merzenich: 1997, 'The Role of Temporal Processing in Developmental Language-Based Learning Disorders: Research and Clinical Implications', in B. Blachman (ed.), *Foundations of Reading Acquisition and Dyslexia: Implications for Early Intervention*, Lawrence Erlbaum Associates, Mahwah, New Jersey, pp. 49–66.

Teller, P.: 2001, 'Twilight of the Perfect Model Model', *Erkenntnis* **55**(3), 393–415

Tooby, J. and L. Cosmides: 1992, 'The Psychological Foundations of Culture', in J. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, Oxford University Press, Oxford.

Wimsatt, W. C.: 1986, 'Forms of Aggregativity', in A. Donagan, N. Perovich and M. Wedin (eds.), *Human Nature and Natural Knowledge*, D. Reidel, Dordrecht.

Department of Philosophy
University of California, Davis
1238 Social Science and Humanities Building
Davis, CA 95616-8673
U.S.A.
rcummins@ucdavis.edu