Forthcoming in *Studies in History and Philosophy of Science*

Classical Computationalism and the Many Problems of Cognitive Relevance

Richard Samuels
Ohio State University

Introduction

This paper is part of a larger project that seeks to defend a general conception of our mental architecture that figured prominently in much early cognitive science.[1] For want of a better name I call it the *Standard Account* (SA). Here it is in capsule form:

> Much human behaviour depends on a general capacity for practical reasoning: a capacity that is subserved by a non-modular classical computational mechanism for means-ends inference –a *Planner*. In keeping with virtually all other approaches, the SA assumes the existence of *input systems* for various kinds of perceptual processing and *output systems* for the production and control of motor behaviour. But in contrast to many views, the SA also assumes the existence of one or more non-modular, classical computational mechanism for theoretical reasoning: mechanisms that revise or update beliefs in the light of new information. These mechanisms interact with the Planner at least to the extent that the latter requires beliefs produced by the former in order to reliably produce successful plans. Although there may be many other cognitive mechanisms, the SA remains silent on such matters. Thus it does not

[1] While I do not claim that this description precisely captures the view of any single theorist, I do maintain it describes a general approach that can be found with varying degrees of explicitness in the work of many, including George Miller's seminal work on the role of planning, Newell and Simon's research on problem solving, and John Anderson's account of cognitive architecture (Miller et al. 1956; Newell & Simon, 1973; Anderson, 1991). In addition, the kind of view depicted by the SA dominated –albeit in unsatisfactory form— much early robotics; and nor is it a million miles from views found in Jerry Fodor's earlier work (Nielson, 1971; Fodor, 1975).

purport to specify a complete cognitive architecture but the kind of *core architecture* represented schematically in Figure 1.

Philosophers and cognitive scientists have challenged almost every aspect of this general picture. In what follows I focus on objections to the thesis that reasoning processes are a species of classical computation. More specifically, I focus on a range of highly influential objections that concern what might loosely be termed *relevance problems* –roughly, problems concerning how to determine which of a range of representations is relevant to the performance of a given cognitive task. According to some critics, the nature of such problems provides reason to reject the classical computational account of reasoning. In what follows, I aim to show that this is not so.

There isn't just one problem of relevance. Rather there are many different problems that give rise to different issues. Historically, these problems are closely associated with the frame problem in artificial intelligence; and (unfortunately) more recent discussions of relevance problems –and their implications for accounts of cognitive architecture— have inherited much of the unclarity that characterized earlier discussions of the frame problem.[2] With this in mind, I proceed by distinguishing, and discussing in turn, a range of different problems. In doing so, I argue that some versions of the relevance problem pose no serious challenge since they depend on false assumptions about the nature of human cognition or about classical computation. In contrast, other versions of the relevance problem are, I maintain, genuine research problems in the sense that a satisfactory computational account of reasoning must address them. But I maintain that there is no reason to suppose that such problems provide a basis for rejecting classical accounts of reasoning.

Here's the game plan. In section 1, I sketch the classical computational account of reasoning. In sections 2, I discuss three problems that are widely associated with concerns over relevance: McCarthy and Hayes' original frame problem, the update problem, and what I call the relevance in update problem. These problems will only be discussed briefly since, as I hope to make clear, it is

[2] Due to the terminological confusions that surround the expression "frame problem", I will reserve that label for the original problem posed by McCarthy and Hayes.

implausible that, as formulated, they constitute grounds for rejecting CCTR. Nevertheless, the discussion of these problems will lead to two further variants of the relevance problem that have been viewed as insurmountable challenges for CCTR: what I'll call the holism problem and the globality problem. In section 3, I consider the holism problem, and argue that there is little reason to suppose that it constitutes the basis for a serious objection to CCTR. Then in section 4, I discuss the globality problem. Though this problem has been raised by a number of theorists, I focus on Fodor's discussion in *The Mind Doesn't Work that Way* since it is, to my knowledge, the most detailed formulation of the globality problem. In response to Fodor's arguments I maintain that they only succeed in undermining a highly implausible version of classicism, but that his arguments against more plausible versions of the thesis fail. I conclude, in section 5, by briefly discussing a question of diagnosis: If relevance problems do not undermine CCTR, then why have efforts to provide plausible classical computational models of reasoning proven so unsuccessful? Contrary to Fodor and others, I defend an epistemic account of the patterns of successes and failures in cognitive science –one that is wholly consistent with the truth of classicism.

Section 1: The Classical Computational Theory of Reasoning

The classical computational theory of reasoning (CCTR) is a restricted version of the classical computational theory of mind. According to this view, reasoning depends on symbol manipulation. Slightly more precisely:

> *CCTR:* Human reasoning processes depend on algorithmically specifiable processes that are causally sensitive to –indeed defined over—the syntactic properties of compositionally structured mental representations.

Though there is, of course, much that could be said about this thesis, it is a very familiar one, and I won't belabor the details here.[3] Still, it would be useful to make clear what this version of CCTR is and is not committed to.

First, the above formulation of CCTR differs from many standard, philosophical formulations of classicism in that it does not purport to provide a metaphysics of mental states and processes (Rey, 1998). For rather than claiming that reasoning *is* a classical computational process, it is instead formulated as a claim

---

[3] I will, in any case, say more about its commitments in section 5.

about the dependence of human reasoning on classical processes. The sort of dependence I have in mind here is nomic dependence. And my reason for formulating the thesis this way is that it comports closely with what I take to be the central explanatory goals of cognitive science –viz. to provide an account of the causal mechanisms which subserve cognition and behavior— without requiring any commitment to what is, in my view, a less plausible claim about the metaphysics of reasoning.

Second, in contrast to some formulations of classical computationalism, the above thesis does not imply any commitment to the claim that reasoning as such –or even different kinds of reasoning— depends upon a single computational process. Of course, one way that a reasoning process could nomologically depend on algorithmic processes is for there to be some algorithm whose implementation is sufficient for the occurrence of reasoning. But reasoning might also depend on the simultaneous action of a considerable number of algorithmic processes, whose interaction is its own simplest description. If this were so, reasoning would still depend on classical computational processes. But to use Marr's well-known distinction: it would a kind of process for which there is no Type 1 theory, only a Type 2 theory (Marr, 1977).

So much for what CCTR's commitments are not. What are it commitments? In brief, here are three commitments:

- *Computability:* First, and most obviously, CCTR is committed to the claim that reasoning processes are computable in roughly the sense that some classical computational device or suite of devices could mimic the input-output relations exhibited by human reasoning processes.
- *Tractable computability:* But CCTR requires more than mere computability. It would be no good, for example, if the only computational accounts of reasoning required that our brains were the size of blimps, or that the computations on which reasoning depends take millennia to perform. The problem with such accounts would be that they are not *tractably* computatable for human beings. The notion of computational tractability has been used in a variety of non-equivalent ways. But for present purposes, the rough idea is that a classical computational account of reasoning should not require of us

more time and resources —memory, information, and computational power— than human beings could reasonably be expected to possess.[4]

- *Psychological plausibility.* CCTR is not a claim about any possible kind of reasoning or about reasoning as such. Rather it is a claim about the causal processes that subserve *human* reasoning. As such, the truth of CCTR requires that there exist some psychologically plausible classical computational account of how we reason –roughly, one that not only captures input-output relations in a tractable manner, but which specifies in a plausible what how we do what we do when reasoning.

It will be useful to keep these commitments in mind since, in what follows, we'll consider a range of 'relevance problems' that have been taken to show that one or other of these commitments cannot be met by the CCTR.

2. Frome Frame Problem to Relevance Problems

Historically, problems of relevance have been closely associated with the frame problem in AI (McCarthy & Hayes, 1969). Indeed almost all of the problems that I discuss have at one time or another been labelled as the frame problem. With

---

[4] According to one characterization of tractability familiar from computer science, an algorithm for solving some problem is tractable if, in the worst case, it is polynomial in the size of the input; that is, the resources required to compute a solution to every input can be expressed as a polynomial (or better) function of input size—e.g., n2 or n300. In contrast, an algorithm is intractable if, in the worst case, it is superpolynomial, in the sense that resource requirements increase exponentially (or worse) as a function of input size and can thus only be expressed as superpolynomial functions, such as 2n or 100n. But for current purposes this characterization of (in)tractability is doubly unsuitable. First, it is very widely assumed on inductive grounds by those who model cognitive processes that pretty much any interesting computational problem is superpolynomial in the worst case. Thus, the current criterion for intractability does little more than characterize those problems that are of interest to a computational account of cognition. Second, it is entirely possible for a superpolynomial algorithm to very frequently— indeed normally— be significantly less expensive than the worst case. In which case, it's hard to see why intractability, in this sense, poses a problem for CCTR. After all, it may just be that performance limitations prevent the algorithm being used in the worst case.

this in mind, I propose to start with the original frame problem and trace a route to those relevance problems that are thought to pose a serious challenge for CCTR.

2.1 The Original Frame Problem

The original frame problem, as formulated by McCarthy & Hayes, is a relatively narrow, technical problem in 'logic-based' approaches to artificial intelligence. In brief, it concerns how to represent those things that *don't* change as a result of performing an action. The problem emerges most directly on a pair of assumptions:

a) A cognitive agent –a robot or human being— has a store of belief-like representations of the world; and
b) This store of representations is updated by a deductive, reasoning process of some sort.

To take a simple example: Suppose that there is a red box in the corner of the living room at time t1; and that a robot then moves the box to the kitchen at time t2. Here's an odd question: What color is the box at t2? The answer is, of course, obvious to us. We assume with little or no need for reflection that the box's color won't change as a consequence of the move. But logic alone does not provide such an answer since it doesn't deductively *follow* from all the facts at t1 plus the fact that a red box is moved, that it remains red at t2.[5] *Mutatis mutandis* for almost any action. What's required is some way of representing facts about non-change so that a robot can deductively infer what does not change as a result of performing a specific action.

How might this be done? One response would be to give the robot lots of representations –so-called 'frame axioms'—that specify what does *not* change as a result of performing an action. For example:

Color(x,c) holds after Move(x,p) if Color(x,c) held beforehand.

But this is obviously unsatisfactory since, for agents with even relatively sophisticated representational capacities –i.e. capable of representing lots of individuals, properties and actions— it would require a huge number of frame axioms.[6] Most of an object's properties simply don't change as a result of performing most specific actions. The problem, then, is to represent in a more succinct manner –i.e. without using frame

[5] The box, for example, might have moved through a paint spray on its way to the kitchen.

[6] Where there are N properties and M actions, there will be almost NM frame axioms.

axioms— the fact that most actions don't change most properties of an object. As it is sometimes put: the frame problem is the problem of representing in a compact manner the fact that our world is (typically) inertial (Shannahan, 1997).

A comment about the character of the original frame problem. Notice that the problem is of a quite specific kind. It is what AI researchers sometimes call a *representational problem*. It is not a problem about how to specify the processes or mechanisms or methods responsible for determining non-change. Rather it is a problem about representing in an appropriate format the fact that our world is inertial. As we will soon see most other problems of relevance are quite different in this regard.

Does the problem undermine CCTR? Though the frame problem is often cited as a serious problem,[7] it is implausible to claim that McCarthy and Hayes' original problem provides us with reason to reject CCTR. First, the claim that human reasoning is, in the general case, a deductive process lacks plausibility. Though there are no doubt many cases in which our reasoning is deductive, there is now widespread consensus that in many instances the logical relationship between earlier and later belief states is a weaker one. Indeed, as we will see in later sections, the claim that much of our reasoning is *not* deductive lies at the very heart of some of the most apparently problematic versions of the relevance problem.

Second, even if belief revision were a deductive process, it is far from clear that the original frame problem would constitute grounds for rejecting CCTR. This is because it seems that computer scientists have found plausible solutions to the relatively narrow technical problem outlined earlier. (See, for example, Shanahan, 1997.) That said, there are problems in the near vicinity of the original frame problem that are thought to pose a genuine challenge to CCTR; and it is to these problems that I now turn.

2.2. The Update Problem

Notice that the context in which the original frame problem arises is one in which a cognitive agent –human or robotic— is faced with the task of determining which of its beliefs to revise and which to leave untouched. This leads fairly directly to what is sometimes called the update problem. In general form the problem is this:

[7] See, for example, Waskam (2003).

> *Generalized Update Problem:* How is an agent to update its beliefs in response to newly available information so as to maintain a suitably veridical belief-system?[8]

Notice that the update problem is not a narrowly representational one in the way that the original frame problem is. Rather it is a *computational problem* that concerns a broad issue about the computational methods and processes involved in belief-revision. Indeed, a complete solution to the update problem would seem to require nothing less than a specification of those computational processes involved in the revision of belief.[9]

Does the problem undermine CCTR? Clearly, the update problem poses a *very* hard research topic for cognitive science. Assuming that CCTR is true, solving the problem is tantamount to providing a comprehensive, computational account of human belief revision; and presumably this is a very large, and very hard topic for research. But without considerable further argument this would not be reason to *reject* CCTR. Rather it is merely the specification of one large and important aspect of the project of cognitive science: viz. to provide a mechanistic account of belief revision. Moreover, this is not merely a problem for research programs that cleave to CCTR. Rather, it is presumably a very hard problem for any research program that does not make the distinctly implausible move of denying that we engage in reasoning. To turn the existence of the update problem into an argument against CCTR, then, one clearly

[8] Notice that the update problem is formulated in terms of the notion of veridical belief. Clearly the task of revising beliefs in just any old way is not a hard one. (Revise randomly!) Rather, it is widely assumed that the problem is to revise beliefs in such a way as to ensure that the resulting representation of the world is a fairly accurate one. One motivation for this is that–at least for mundane matters— it really does seem that our beliefs *are* fairly accurate. Another is that the SA presupposes a reasonably accurate representation of the world. Nevertheless, it would be wrong to assume some overly demanding standard –e.g. that beliefs must be wholly accurate. Such a standard ought not to be imposed, in part, because human beings clearly fail to meet such a standard. Rather, to sue Dennett's memorable phrase, our beliefs no only be good enough for government work.

[9] Historical Note: The update problem is a close analogue of what Hayes called the Generalized AI Problem –or GAIP (Hayes, 1987).

needs to provide some plausible grounds for supposing that the problem is insoluble within a classical computational approach.

Why is the update problem such a challenging research issue? There are many reasons. But one that leads us in the direction of relevance problems is the representational *richness* of the sorts of cognitive systems we seek to understand. Suppose, contrary to fact, that we were only interested in systems that are representationally impoverished in the sense that they possess only a limited number of belief-like representations. Such systems could determine which representations to revise in the light of new information by brute force: by considering all of their representational states. But human beings are not plausibly viewed as representationally impoverished in this way. Rather we appear to possess a *huge* number of belief-like representations (as well as many other kinds of representation). For systems like us, then, it is implausible to suppose that *all* beliefs get considered. Amongst other things, such a proposal would risk being computationally intractable. For such a system, then, it will be necessary to determine those beliefs that are *relevant* to any particular update task.[10]

2.3 Relevance Problems

This leads us to *relevance problems*. Roughly put, such problems conform to the following general schema:

> *Relevance Problems*: Given a task, T, and computational system S, how does S determine (with reasonable levels of success) from all the available information which is *relevant* to the specific task at hand? (Glymour, 1987)

Such problems can arise in the performance of many different tasks –planning, decision-making, pragmatics, perception, and so on. But in the present context, it is

[10] Of course, this does not mean that the system must consider *all and only* those representations that are relevant to a particular task. There is little reason to suppose that humans exhibit any such precise sensitivity to relevance. (Indeed, absent some precise account of what relevance is, it's unclear what sense can be given to this idea.) What seems more appropriate as a criterion is that of the beliefs we consider in the light of new information, some significant subset are typically relevant to the task of updating one's belief system.

belief revision tasks that are of most immediate interest. That is, we are concerned primarily with the *problem of relevance in update*:

> *Relevance in Update*: Given some new information, how do we determine (with reasonable levels of success) which of the representational states we possess are *relevant* to determining how to update our beliefs?

Does the problem undermine CCTR? As with the Update Problem, the present problem is presumably a very hard research topic for cognitive science. Amongst other things, it requires the specification of tractable, psychologically plausible, computational processes that manage –without considering all beliefs— to use those beliefs that are relevant to a the revision of any given belief. But again, the fact that the problem constitutes a hard research topic is not, by itself, reason to reject CCTR. Rather, it is merely the specification of one central part of the problem of explaining belief revision. What is required to turn this into an *objection* to CCTR is some plausible elaboration of the problem on which it is implausible that CCTR can accommodate the sort of relevance-sensitivity characteristic of human reasoning. In what follows, I consider perhaps the two most influential elaborations of this sort. The first gives rise to what has sometimes been called the problem of holism (Lormand, 1994). The second gives rise to what has been called the globality problem (Fodor, 2000).

3. The Problem of Holism

The holism problem arises from the apparent fact that much human reasoning is holistic. The sort of holism at issue is *not* that all –or even most—of our beliefs *actually* figure in any specific instance of reasoning. (As we have already seen, it is precisely the fact that not all beliefs get used when reasoning that raises the relevance in update problem in the first place.) Instead, the sort of holism at stake here is modal in character. What it amounts to is that under the appropriate conditions –especially those involving different background beliefs— the relevance of a belief to a reasoning task *can* vary dramatically. Slightly more precisely:

Inferential Holism: Given appropriate background beliefs, (almost) any belief can be rendered relevant to the assessment of (almost)[11] any other belief.

Or to use Fodor (1983)'s terminology: belief revision processes are *isotropic*. To take a fairly simple example[12]: On the face of it, the current cost of tea in China has little to do with whether or not my brother's baby in England will cry on Saturday morning. But suppose that I believe my brother has stocks invested in Chinese tea, that he reads the business section of the newspaper every Saturday morning, and that on reading bad financial news he tends to fly into a rage. Given these background beliefs, it seems that beliefs about the current cost of tea in China may well be relevant to beliefs about whether or not my brother's baby will cry on Saturday morning.[13] *Mutatis mutandis* for other beliefs. Or so it would seem. In which case, we appear to have the following problem:

> *Holism Problem:* Given that (almost) any belief can be relevant, under the appropriate conditions, to the assessment of (almost) any other, how do we determine (with reasonable levels of success) which of our beliefs are in fact *relevant* to a specific instance of belief revision?

3.2. Does the problem undermine CCTR?

Does the holism problem provide a reason to reject CCTR? If belief revision is holistic, then it clearly lies at the heart of a very hard topic for research –viz. the explanation of belief revision. Moreover, the problem of holism seems to *intensify* the relevance in update problem (Lormand, 1994). One reason for this is that holism precludes an obvious solution to the task of determining which representations are relevant to a given instance of belief revision. Specifically, one might have thought that it was possible for beliefs to be partitioned in advance into particular categories according to which are relevant to what (types of) hypotheses –e.g. that one set of beliefs was relevant to biological hypotheses, another to economic hypotheses, another to psychological hypotheses, as so on. If this were possible, then one might

---

[11] Clearly, this could do with refinement. So, for example, few beliefs will presumably be relevant to the assessment of logical beliefs –e.g. that if P, then P.

[12] The example is, I think, based on a case used in Copeland (1993) who, in turn, was based on an example from Guha and Levy (1990).

[13] Perhaps I should also add that children find angry fathers scary, and that they cry when afraid.

think that a reasoning system, on performing a specific kind of inferential task, could easily access those beliefs that are relevant, and safely ignore the rest. In which case the relevance in update problem would be solved. But if human reasoning is holistic, then there could be no such clear *a priori* division. For such a proposal would fail to accommodate the fact that alternations in one's background beliefs can make any belief relevant to the assessment of any other.

So, the holism problem is plausibly a serious research challenge for CCRT. But why suppose that this provides us with reason to *reject* CCRT as opposed to merely being a hard research topic? I know of two arguments.

3.2.1: Argument 1: Optimal Relevance Sensitivity. The first argument, due to Dietrich & Fields (1996), was primarily intended to undermine Fodor's claim that central processes are isotropic. But since Fodor means by "isotropy" what I mean here by "holism", the argument would, if sound, show the impossibility of holistic computational systems.

The argument seeks to draw out the putatively implausible implications of the assumption that there are holistic cognitive processes. The first premise provides a gloss of what genuine holism demands:

1. A genuinely holistic process never considers an arbitrary subset of the information relevant to confirming a hypothesis.

The second premise specifies a typical condition on determining the relevance of a piece of information.

2. The relevance of a piece of information can rarely be determined in advance.

In other words, determining the relevance of a piece of information to an hypothesis typically requires that one actually assess the degree to which the information (dis)confirms the hypothesis. But if premises 1 and 2 are both true, then according to Dietrich and Field:

3. For a computational system is to be genuinely holistic, it must access all of the available information just to confirm a single hypothesis.

Indeed, according to Dietrich and Field, a genuinely holistic system would not merely need to consider all the information available to the agent, but "all the information in the Universe"! (Dietrich & Field, 1996). Unsurprisingly, they conclude that a genuinely holistic process would be unfeasible.

There are, however, some problems with the argument when construed as an objection to CCTR. Perhaps the most obvious is that the notion of holism that it employs is too strong. On Dietrich and Fields' view, a holistic process always considers all the relevant information. But the notion of holism in play here –and the Fodorian notion of isotropy, for that matter— makes no such commitment. Rather, as noted earlier, the appropriate notion of holism is a modal one. A process would be holistic if, given the appropriate background beliefs, it *could* bring (almost) any belief to bear on the assessment of (almost) any hypothesis. But this is clearly much weaker than Dietrich and Field's conception of holism. For one thing, it is wholly consistent with the idea that holistic processes very often fail to consider all the relevant information, and indeed that they also consider information that is irrelevant.

3.2.2. Argument 2: The intractability of unencapsulated (non-modular) processes. The second argument from holism to the rejection of CCTR is closely related to some familiar arguments against the existence of unencapsulated reasoning processes (Samuels, 2005). In brief, an encapsulated cognitive process is one that ''has access, in the course of its computations, to less than all of the information at the disposal of the organism whose cognitive faculty it is" (Fodor, 1987, p. 25). Paradigmatic examples—such as those involved in length perception or phonological processing—cannot draw upon the full range of the organism's beliefs. In contrast, a process that can access (virtually) all of our beliefs would be the paradigm of a highly unencapsulated cognitive process (Fodor, 1983; Stanovich, 1999).[14] With this in mind, the argument can be presented as follows:

1. If belief revision is holistic –if it can, in principle, bring any belief to bear on the assessment of a hypothesis— then it is unencapsulated.
2. For a computational process to be tractable, it must be encapsulated.
3. So: If belief revision is holistic, it is computationally intractable.

What are we to make of this argument? The first premise is plausible. If almost any belief can, given appropriate background beliefs, be brought to bear on the assessment of almost any other, then we have reason to think that belief revision is an

[14] For further detail on the notion of encapsulation see Fodor (2000), Samuels (2005) and Carruthers (2006).

unencapsulated process. The problem is with the second premise. There is a long story to tell here (See Samuels, 2005). But the short version is that tractability does not require encapsulation. As with most real-world computational applications –web search engines, for example— there may be heuristic and approximation techniques that permit feasible computation: techniques that often, though not invariably, identify a substantial subset of those representations that are relevant to the task at hand. Of course, this would not be an option if one maintained that, when reasoning, we are *guaranteed* to identify relevant beliefs. But I can see no reason whatsoever to suppose that this claim is true. Indeed, one very clear moral of the last four decades of research on human judgment and decision-making is that such standards of accuracy are misplaced.[15]

4. The Globality Problem

Let's turn to the globality problem. Though the worry has been discussed by a number of recent theorists, I focus primarily on Fodor's influential formulation of the problem (Fodor, 2000; Fodor 2005. See also Sperber, 2005 and Carruthers, 2006). In roughest outline, Fodor maintains that the difficulties we have addressing relevance problems within a classical framework have a common root cause: the globality of reasoning. Slightly more precisely, he maintains that there is a fundamental tension between the local, syntactically determined character of classical computation and the global character of much human reasoning, especially abductive inference and planning. In what follows, I first discuss the claim that reasoning is global (4.1) and the claim that classical computations are local, syntactically determined ones (4.2.). I then set out and respond to Fodor's arguments for the conclusion that globality undermines CCTR (4.3 and 4.4). I show that Fodor's arguments only succeed in showing that an independently implausible version of classical computationalism is false.

---

[15] It's also worth noting that if the problem of holism is a genuine problem –if human reasoning is holistic in the present modal sense—then it is a problem not merely for CCTR but for any conception of reasoning. For example, a connectionist theory must also explain this fact. And vague talk of building relevance in the connections between nodes (Horgan & Tienson, 1996) simply won't do.

4.1. What is a global process?

According to Fodor an important class of reasoning processes –which includes abduction and planning— are *global* processes in the sense that they are casually sensitive to global properties of representations. Though it is not entirely clear from Fodor's discussion what a global property is, the following conditions appear most central:

- *Context variability:* A global property of a representation, R, is not one that R has invariably. Rather it is a property that it can have or fail to have depending on context.
- *Dependence on doxastic context:* A global property of a representation, R, is sensitive not just to any contextual factor but rather to the doxastic context in which it is embedded. In particular, whether R possesses a global property, G, will vary as a function of changes in the agent's background beliefs.
- *Large Contexts:* Satisfying the above two conditions would make a property, of a representation sensitive to doxastic context. But in order that a property G be a global property a further condition needs to be satisfied. The aspects of the doxastic context that are relevant to determining whether R has G must be very *large:* paradigmatically the entire belief system, or at any rate some very substantial part thereof.

By way of illustration, consider abductive inferences of the sort that figure in scientific enquiry. According to Fodor, abduction is sensitive to global properties of the representations involved. In particular, when selecting between completing beliefs or hypotheses, we often assess their simplicity and conservativism: properties that Fodor takes to paradigmatic examples of global properties. Consider the property of simplicity. A belief, B, can be more or less simple depending on what background beliefs you have.[16] Moreover, according to Fodor, B's simplicity will not depend on some small, isolable component of your belief system, but typically on some large part of it –whole theories or even the totality of one's epistemic commitments (Fodor, 2000, p. 33). But if this is so, then abductive reasoning is a global process.

What does any of this have to do with problems of relevance? Like the holism problem, globality intensifies the problem of relevance. Holism, you may recall, appears to imply that the beliefs relevant to a given inferential task can, in principle,

---

[16] Mutatis mutandis hypotheses, assumptions and plans.

come from anywhere in one's belief system. The globality problem goes one step further. As we will see, what Fodor takes the problem to show is that for almost any instance of belief revision, the set of beliefs that a classical computational device would need to consider will be very large indeed: whole theories or even the totality of one's epistemic commitments (Fodor, 2000). In other words, it's not merely that relevant beliefs might come from anywhere, it's that they routinely come from (almost) *everywhere*. The relevance problem with avengeance.

4.2 Syntactic Determination

So, according to Fodor, some reasoning processes –most obviously abduction—are global processes. What of the claim that computational processes of the sort invoked by CCTR are local, syntactically determined ones? As Fodor notes, there is a systematic ambiguity in talk of syntactic determination. According to one reading:

E(CTM): The causal role of a mental representation, R, in cognitive processes is determined by its essential syntactic properties –i.e. *its constituent structure.*

Take a mental representation –e.g. THE CAT IS ON THE MAT. What E(CTM) implies is that its causal role in cognitive processes is wholly determined by the representations from which it is constituted –THE, CAT, IS etc.—plus their mode of combination. It's worth stressing that this has some surprisingly strong consequences.[17] First, it seems to imply that only essential properties determine the causal role of a representation. This is because, as Fodor notes, the constituent structure of any syntactically structured representation is an essential property of that representation. (Roughly: Necessarily, if two representations differ in their syntactic structure, then they are thereby not the same representation.) Moreover, since an entities essential properties cannot not vary from context to content, it follows from the discussion so far that the causal role of a representation is solely determined be its *context invariant* properties.[18] As I said this really is a very strong claim. But E(CTM)

[17] I'm tempted to think that the "E" in "E(CTM)" strands for "extreme".

[18] Some of Fodor's comments also suggest that he thinks of constituent syntactic properties as intrinsic properties. For the record, I find this assumption quite implausible. But since the assumption plays no role in the coming discussion, I omit it from the main text.

is not the only construal of syntactic determination An alternative and more moderate reading is this:

> M(CTM): The causal role of a mental representation, R, in cognitive processes supervenes on *some syntactic properties or other.*

This is far weaker than E(CTM) in that it allows that R's causal role depend not only by its constituent structure, but also on its inessential syntactic properties–e.g. it's syntactic relations to other representations— and on the syntactic properties of representations other than R. Fodor claims that whichever conception of syntactic determination one adopts, CCTR is in serious trouble.

4.3 The Objection to E(CTM)

Given the above, we are in a position to see that E(CTM) is implausible as a claim about cognition in general. First suppose that:

1) E(CTM) is true.

Then it follows from the definition of E(CTM) that:

2) The causal role of any representation R is wholly determined by R's constituent structure.

Further, given the earlier claims about constituent structure of representations, it follows from 2) that:

3) R's causal role is wholly determined by R's essential, hence, context invariant properties.

But 3) is implausible in the case of abduction (and planning). For in such cases, R's causal role depends, in part, on context variable, hence, inessential properties of R – such as R's simplicity. In which case:

4) R's causal role is not wholly determined by R's essential, hence, context invariant properties.

Clearly, 3) and 4) are inconsistent with each other. In which case, we have reason to suppose that the original supposition, E(CTM), is false.

Some comments on the argument are in order. First, the above is, in my view, a charitable reconstruction of Fodor's argument against E(CTM).[19] Moreover, as far

[19] See Ludwig & Schneider (2008) for a different and, in my view, rather less charitable reconstruction.

as I can tell, the argument is a sound one that shows E(CTM) to be false. But second, it is important to stress that the argument has little to do with the *globality* of abduction. What matters to the argument is that abduction depends on *context variant* properties of representations, not that these properties are global. To be sure, the causal relevance of global properties implies the causal relevance of context sensitive properties. But the converse is not true. For even if simplicity were context sensitive but not global, the falsity of E(CTM) would still follow. Thus we can readily endorse Fodor's argument against E(CTM) without insisting that global properties are even partially determinative of a representation's causal role.

Finally, I maintain that whilst Fodor succeeds in showing that E(CTM) is false, this poses no real threat to CCTR. Indeed I maintain that it should be *obvious* that the falsity of E(CTM) poses no real threat to CCTR. This is because the sorts of proposals that classical computationalists routinely endorse are clearly incompatible with E(CTM). This has nothing to do with globality. Rather, it is just that in paradigmatic examples of classical computational systems factors other than the constituent structure of a mental representation help determine its causal role. For example:

- *Programs:* If the program executed by a computational system is altered, then the causal role of a given representation may well change. (Example: the program is for proving theorems in modal logic. If we change the program from an S4 to S5 theorem prover, then $\Box$P will have a different causal role.)
- *Rules and Heuristics:* If a rule or heuristic is added to a computational system, then the system may well treat representations differently than it previously did. Hence the causal role of the representation will change. (Example: Modify a chess program by adding new heuristics for determining a good move.)
- *Items in memory:* Even where no change is made to the program, rules or heuristics for a classical system, the mere addition of new representations to a database will typically change what inferences a representation can participate in.

In each of these cases, we have a difference in the causal role of a representation without a difference in its constituent structure. But all these sorts of changes are routine in the sorts of systems that have traditionally been construed as classical ones. I conclude, then, that whilst E(CTM) is false, this should be of no concern whatsoever to classical cognitive scientists.

4.4 Objections to M(CTM)

Let's now turn to the more moderate M(CTM). According to this thesis, the causal role of a mental representation in cognitive processes supervenes on some syntactic fact or other. As Fodor notes, this is wholly compatible with the previous objection to E(CTM). But even so, he maintains that M(CTM) is unsatisfactory. He offers two main arguments in support of this claim.

4.4.1 The Units of Cognition Argument

According to Fodor, M(CTM) is psychologically implausible because it commits us an implausible view about the size of the units of cognition –i.e. those things we confirm, entertain, assess, and so on. The argument proceeds via two observations. First, on the assumption that M(CTM) is true, a reasoning process could only be sensitive to a global property of a representation, R, if that property supervened on the constituent structure of some set of representations, K. Second, the primitive operations of a classical computer are local in the sense that they are defined over –and apply to— the primitive vocabulary of the language that the machine computes in –e.g. writing a symbol, deleting a symbol etc. Given these assumptions it would seem to follow that:

1. For a classical device to be sensitive to a global property –one that supervenes not merely by R's constituent structure but on the constituent structure of other representations— it must perform primitive operations not merely to R but to R and members of K.

In some cases, this might not be a problem. But where a classical system is sensitive to global properties, it would seem that K must be very large. That is:

2. For global processes, such as abduction, K will be *very* large. Indeed K will typically include whole theories or even the entire belief system.

According to Fodor, however, this is incompatible with plausible assumptions about what the units of cognition –of confirmation, in particular—typically are.

3. The units of cognition are typically far smaller than whole theories. Indeed "the typical unit of confirmation is a judgment that an individual has a certain property" (Fodor, p.32).

In which case, it would seem to follow that:

4. A commitment to M(CTM) results in a psychologically implausible account of cognition: one on which the typical units of confirmation would be far larger than they in fact are.

Criticism. The argument does not work. One relatively minor worry is that in developing the argument Fodor assumes the falsity of confirmational holism with little or no argument. But even if one assumes the falsity of confirmational holism, the argument still fails. This is because it collapses a distinction between:

- The units of confirmation: roughly, that which gets confirmed or disconfirmed; and
- Those considerations are relevant to assessing or confirming something.[20]

At most what the argument supports is the conclusion that lots of beliefs –i.e. K— will need to be accessed by a classical computational device in order that it be sensitive to the global properties of a representation. But that doesn't show that K is part of the unit of confirmation –i.e. that which gets (dis)confirmed. All it shows is that K is among those things relevant to the (dis)confirmation of R. In short: the claim that individual judgments are the typical units confirmation seems wholly consistent with the (putative) fact that we access a great many beliefs in the course of (dis)confirming a judgment.

3.4.2 The Tractability Argument

A second kind of argument that has been widely attributed to Fodor is that a global but computational process –e.g. for abduction— would be computationally intractable. According to Fodor, for a classical computational process to be sensitive to a global property of a representation, R, it would require access not only to R but also to all those other representations on which the property depends. For example, if the simplicity and conservativism of a hypothesis, H, depend not only on *its* constituent structure, but also one's background beliefs, K, then a computational process that is sensitive to the simplicity of H must also have access to elements of K. Given this putative fact, the obvious question is: How *much* of K needs to be consulted in order for a classical system to perform reliable abduction? If it were only

---

[20] Oddly, this is a distinction that Fodor is on occasion keen to draw. (See, for example, Fodor, 2000, chap. 2)

a handful of other beliefs, then no obvious problems for CCTR follow. But according Fodor, it will typically be the case that *lots* of an agent's background beliefs will need to be accessed in the course of abductive reasoning. Indeed, he appears to think that the totality of K will very often need to be accessed since this is the "only guaranteed way" of classically computing a global property. But this surely would render reliable abduction computationally intractable:

> 'Reliable abduction may require, in the limit, that the whole background of epistemic commitments be somehow brought to bear on planning and belief fixation. But feasible abduction requires in practice that not more than a small subset of even the relevant background beliefs are actually consulted.' (2000, p. 37)

In which case, it's hard to see how CCTR could be right about abductive reasoning.

Criticism. Although inferences involving simplicity and conservativism are plausibly context dependent, Fodor provides us with no reason whatsoever to think that they are global in any sense that threatens CCTR. In order to see this, it is important to keep firmly in mind the general distinction between normative and descriptive-psychological claims about reasoning: claims about how we ought to reason and claims about how we actually reason. This distinction applies to the specific case of assessing the simplicity and conservativism of hypotheses. On the normative reading, assessments of simplicity and conservativism ought to be global: that is, normatively correct assessments ought to take into consideration one's total background epistemic commitments. But of course it is not enough for Fodor's purposes that such assessments ought to be global.[21] Rather, it needs to be the case that the assessments humans make are, in fact, global; and to my knowledge, there is no reason whatsoever to suppose that this is true. So, whilst it may well be the case that we assess beliefs for their simplicity and conservativism, it's far from clear that this means that cognitive processes are sensitive to the global properties of beliefs.

A comparison with the notion of consistency may help to make the point clearer. Consistency is frequently construed as a normative standard against which to assess one's beliefs (Dennett, 1987). Roughly: all else being equal, one's beliefs

[21] It's also far from clear that a normative theory of reasoning mandates that we *ought* to assess the simplicity and conservativism of hypotheses, in so far as they are global properties of beliefs. I leave this issue to one side.

ought to be consistent with each other. When construed in this manner, however, it is natural to think that consistency should be a global property in the sense that any belief ought to be consistent with the entirety of one's background beliefs. But there is absolutely no reason to suppose—and indeed some reason to deny—that human beings conform to this norm (Cherniak, 1986). Moreover, this is so in spite of the fact that consistency really does play a role in our inferential practices. For we really do, on occasion, assess the consistency of some (small) subset of our beliefs. What I am suggesting is that much the same may be true of simplicity and conservativism. When construed in a normative manner, it is natural[22] to think of them as global properties. But when construed as properties of the beliefs that figure in actual human inference, there is no reason to suppose that they accord with this normative characterization.

Second, even if we suppose that the simplicity and conservativism are global properties of actual beliefs, the argument still does not go through, since it turns on the implausible assumption that we are guaranteed to make successful assessments of simplicity and conservativism. Specifically, in arguing for the conclusion that abduction is computationally unfeasible, Fodor relies on the claim that ''the only guaranteed way of Classically computing a syntactic-but-global property'' is to take ''whole theories as computational domains'' (2000, p. 36). But guarantees are *beside the point*. Why suppose that we always successfully compute the global properties on which abduction depends? Presumably we do not. And one very plausible suggestion is that we fail to do so when the cognitive demands required are just too great. In particular, for all that is known, we may well fail under precisely those circumstances that the classical view would predict—namely, when too much of a belief system needs to be consulted in order to compute the simplicity or conservativism of a given belief.

5. Explaining Patterns of Success and Failure

So far I have considered a range of relevance problems and argued that they provide us with no good reason to reject CCTR. Some of these putative problems are ones that we do not solve; some are, for all the arguments show, no more than hard research problems; and some are problems only on very implausible views about the commitments of CCTR. I close by very briefly considering a question of diagnosis:

[22] Though by no means mandatory.

Why has classical cognitive science performed so poorly in trying to provide computational accounts of reasoning?

According to Fodor and others, the failures of classical cognitive science are intimately bound up with the problems that globality and relevance pose. In particular, the failures of cognitive science are most striking for global processes – such as planning and abduction—where, in contrast to other regions of psychology, our efforts to produce computational models has met with very limited success.[23] Moreover, Fodor claims that there is a pattern to these failures:

> Because of the context sensitivity of many parameters of quotidian abductive inference there is no way to delimit a priori what may be relevant to assessing them. In fact there is a familiar dilemma: Reliable abduction may require, in the limit, that the whole background of epistemic commitments be somehow brought to bear on planning and belief fixation. But feasible abduction requires in practice that not more than a small subset of even the relevant background beliefs are actually consulted.' (2000, p. 37)

In short: "it is *because* of the frame problem that our robots don't work"; and it is because of the globality problem that the frame problem appears insoluble (Fodor, p.38). Our failures to classically model abduction are thus traceable to the fact that classical computation does not have the resources to explain such processes.

Diagnosing patterns of success and failure is not a simple affair. Nevertheless, it seems that there is an alternative to Fodor's proposal that is no less plausible as a line of explanation and wholly consistent with CCTR. In brief: the difficulties that we encounter in trying to understand reasoning are principally a consequence of some serious epistemic difficulties that we should expect to confront in constructing theories of reasoning.

In my view, perhaps the most serious of these epistemic difficulties arises from the fact that reasoning processes are 'hidden' between perceptual systems and systems for motor control. One consequence of this is that it's very hard to specify many of the properties that are crucial to understanding a cognitive system. For example, in order to provide a computational account of a cognitive process, one needs to specify the function computed. But it is not enough to specify the inputs and outputs

[23] Indeed, he has suggested that the purpose of his discussion of globality was never to argue against classicism but to provide a diagnosis of why it fails. See Fodor (2005).

intentionally. Rather, one needs to provide an account of the causally relevant non-semantic –i.e. syntactic— properties of the representations involved. The problem is that this is very hard to do for non-peripheral processes since we have no readily available means of acquiring evidence about such things. In the case of perceptual processes there has been a very effective strategy of "working in from the outside". In the case of vision, for example, we can acquire first acquire an understanding the properties of retinal inputs, and the use this as a constraint on theories of edge detection, which in turn can be used as a constraint on accounts of other perceptual processes. Similar points apply to the development of theories of motor control. But in the case of reasoning, such a strategy is unavailable. As a consequence, testing and constraining mechanistic hypotheses about reasoning proves very hard indeed.

Clearly, there is much more to say about why reasoning might be such a hard phenomenon to theorize about. But the above consideration is, I think, important and plausibly explains a number of obvious features of the pattern of successes and failures in cognitive science. First it helps explain why in classical cognitive science, computational models of perceptual processes and motor control are far better developed than models of human reasoning. Second, it helps explain why much the same pattern is replicated across all other research traditions. For example, connectionist accounts of perceptual processes and motor control are also far better developed than connectionist accounts of reasoning. Finally, the above observation helps explain why human reasoning processes that are apparently *not* sensitive to global properties are also relatively poorly understood. Specifically, despite decades of enquiry, explicit computational models of deductive reasoning in humans are still woefully underspecified.

To summarize: Even if reasoning processes are not global in Fodor's sense – indeed, even if CCTR is true—we would still have ample grounds to expect the task of computationally modelling reasoning to be very hard indeed.

Bibliography

Carruthers, P. (2003), On Fodor's Problem, *Mind and Language*, vol. 18(5), pp. 502-523.

Cherniak, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.

Cosmides, L. and Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In Hirschfeld and Gelman (1994) *Mapping the Mind*, pp. 85-116.

Fodor, J. (1975).*The Language of Thought.* Harvard University Press.

Fodor, J. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.

Fodor, J. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. Cambridge, MA: MIT Press.

Fodor, L. (2005). "Reply to Steven Pinker 'So How *Does* The Mind Work?'" *Mind & Language* Vol. 20, No. 1, pp. 25-32.

Ford, K.M. & Pylyshyn, Z.W. (eds.) (1996), *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, Norwood, NJ: Ablex.

Guha, R.V. and A.Levy (1990) "A Relevance Based Meta Level" *MCC technical report number ACT-CYC-040-90*.

Hayes, Patrick (1987). "What the frame problem is and isn't." In Z. Pylyshyn (ed.), *The Robot's Dilemma*. Ablex.

Horgan, T. & J. Tienson (1996) *Connectionism and the Philosophy of Psychology*, MIT Press

Lormand, Eric (1994). "The holorobophobe's dilemma." In K. Ford & Z. Pylylshyn (eds.), *The Robot's Dilemma Revisited*. Ablex.

Ludwig, K. & Schneider, S (2007). "Fodor's Challenge to the Classical Computational Theory of Mind" *Mind & Language*, Vol. 23, No. 1., pp. 123-143.

Marr, D. (1981) "Artificial intelligence: a personal view" In Haugeland, J., (ed.) *Mind Design*, chapter 4, pages 129-142. MIT Press, Cambridge, MA.

McCarthy, J. & Hayes, P.J. (1969), "Some Philosophical Problems from the Standpoint of Artificial Intelligence", in *Machine Intelligence 4*, ed. D.Michie and B.Meltzer, Edinburgh: Edinburgh University Press, pp. 463-502.

Pylyshyn, Z. (1987) *The Robots Dilemma: The Frame Problem in Artificial Intelligence (Theoretical Issues in Cognitive Science),* Ablex

Samuels, R. (2005). The complexity of cognition: tractability arguments for massive modularity. In P. Carruthers, S. Laurence, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 107-121). Oxford, England: Oxford University Press.

Samuels, R. (2006). Is the mind massively modular? In R. Stainton (ed.), *Contemporary Debates in Cognitive Science* (pp. 37-56). Oxford, UK: Blackwell.

Shanahan, M.P. (1997), *Solving the Frame Problem: A Mathematical Investigation of the Common Sense Law of Inertia*, Cambridge, MA: MIT Press.

Sperber, D. & Wilson, D. (1996), "Fodor's Frame Problem and Relevance Theory", *Behavioral and Brain Sciences*, vol. 19(3), pp. 530-532.

Stanovich, K. 2004. *The Robot's Rebellion*: *Finding Meaning in the Age of Darwin*. Chicago, Il.: The University of Chicago Press.

Waskan, Jonathan A. (2003). "Intrinsic Cognitive Models." *Cognitive Science*, Vol. 27, No. 2, pp. 259-283.

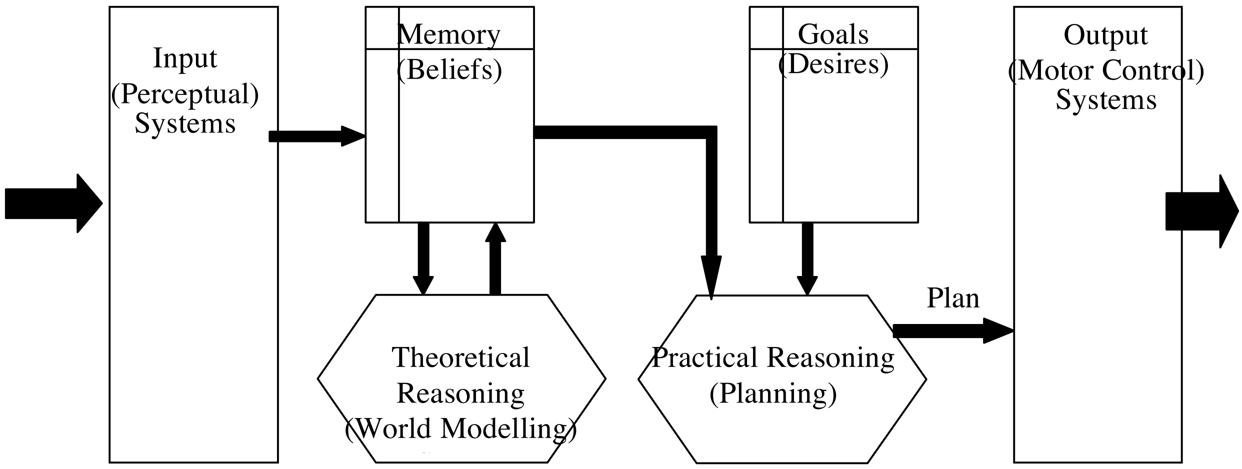
Input
(Perceptual)
Systems
Memory
(Beliefs)
Goals
(Desires)
Output
(Motor Control)
Systems
Theoretical
Reasoning
(World Modelling)
Practical Reasoning
(Planning)
Plan

FIGURE 1