in R, J. Stainton, M. Ezcurdia, and C. D. Viger (eds.)
New Essays in Philosophy of Language and Mind.
Supplementary issue 30 of the Canadian Journal of Philosophy

What Computations (Still, Still) Can't Do: Jerry Fodor on Computation and Modularity

Robert A. Wilson
Department of Philosophy
University of Alberta

Abstract

Fodor's thinking on modularity has been influential throughout a range of the areas studying cognition, chiefly as a prod for positive work on modularity and domain-specificity. In *The Mind Doesn't Work That Way*, Fodor has developed the dark message of *The Modularity of Mind* regarding the limits to modularity and computational analyses. This paper offers a critical assessment of Fodor's scepticism with an eye to highlighting some broader issues in play, including the nature of computation and the role of recent empirical developments in the cognitive sciences in assessing Fodor's position.

in R, J. Stainton, M. Ezcurdia, & C. D. Viger (eds.)
New Essays in Philosophy of Language and Mind.
Supplementary issue 30 of the *Canadian Journal of Philosophy*

What Computations (Still, Still) Can't Do: Jerry Fodor on Computation and Modularity*

1. Introduction

Jerry Fodor's *The Mind Doesn't Work That Way* (2000; hereafter *Mind*) purports to do a number of things. To name three: First, it aims to show what is problematic about recent evolutionary psychology, especially as popularized in Steven Pinker's *How the Mind Works* (1997). Fodor's particular target here is the rose-colored view of evolutionary psychology as offering a "new synthesis" in integrating computational psychology with evolutionary theory. Second, it poses a series of related, in principle problems for any cognitive theory that revolve around the putative tension between the *local* nature of computational processing and the *global* nature of at least some cognitive processing. And third, it reiterates Fodor's earlier argument, in *The Modularity of Mind*, for the hopelessness of trying to extend the notion of modularity from "input systems" to "central systems".

The third of these themes is developed via the second, which in turn provides the basis for the first. I shall concentrate here on the more fundamental parts to this overall argument, namely, the second and third claims listed above. Even though the chief aim of *Mind* is to deflate the current enthusiasm for evolutionary psychology in the cognitive science community, what I find most interesting about Fodor's argument is the way in which it draws on familiar and widely accepted views—about the

---

* A skimpier version of this paper was initially presented as an invited talk at a symposium on Language, Abduction, and Computation at the Canadian Philosophical Association in Quebec City, Canada, in May 2001. I thank Rob Stainton and Chris Viger for the invitation to participate, and the audience for some critical feedback.

nature of computation and cognition—in doing so. So I shall not really discuss Fodor's critique of evolutionary psychology here (see Okasha 2003), but will examine views that underlie that critique.

Part of my interest here stems from my view that such views are in fact mistaken. It is no part of my aim to defend evolutionary psychology, here or elsewhere. Instead, I take reflection on Fodor's critique of evolutionary psychology to provide an opportunity to tackle several views that have become dogmata in the cognitive sciences.

2. Some Cognitive Dissonance for an Aficionado of Fodoriana

Let me begin confessionally: I am puzzled. And although I'm prepared to adjust my set, I am sort of hoping that the fault is in reality. Here is the puzzle. For someone like me, raised on Fodor's earlier writings and the debates they gave shape to--from *The Language of Thought* to the essays in *Representations* to *The Modularity of Mind*—Fodor has been the leading figure in the naturalistic turn in the philosophy of mind that started in the 1960s. Fodor taught many of us, by example, to take psychology and the cognitive sciences seriously in our philosophical thinking about the mind. But when we look to Fodor's work since that time (from *Psychosemantics* to *A Theory of Content and Other Essays* to *The Elm and the Expert* to *Concepts: Where Cognitive Science Went Wrong* to *In Critical Condition* to *Mind*), the naturalistic turn seems to have become, well … the crotchety turn. "Taking psychology seriously" for Fodor for the last 20 years has amounted chiefly to showing that major theoretical departures within the cognitive sciences from the views that Fodor first articulated and defended prior to that time—the connectionist modeling of the 1980s, the techniques of cognitive neuroscience and the neural turn of the 1990s, and now the extension of modularity theory within evolutionary psychology—are all fundamentally mistaken. One might expect that the basis for rejecting these views has been the resounding empirical success of their competitors, or there own empirical shortfalls, but for the most part this hasn't been Fodor's argumentative tack. Rather, it has been strangely *a priori*. The puzzle, in short, is that Fodor's work since *Modularity* hasn't practiced what the earlier work both practiced and preached. What has happened?

Something radical, to be sure. But we can sneak up on this, and lessen the shock to the system, by taking the puzzle in the small at first. What has happened to Fodor's views of modularity and computation in the last 20 years?

In *Modularity*, Fodor proposed a bold, two part thesis about the structure of the mind: that part of it was *modular*—roughly speaking, what he called the "input systems"—and part of it was *non-modular*—what he called "central systems". Fodor argued that we have a real chance of understanding only the modular part of the mind using the resources of computational psychology. Moreover, the reason why we'd never understand the non-modular part of cognition was that it involved processes that were sensitive, in several ways, to properties of the entire cognitive system, "global properties". Hence Fodor's First Law for the Non-Existence of Cognitive Science: the more global a psychological process, the less chance anyone has of understanding it.

Fodor's two part thesis, particularly the first part, spurred a whole range of research into modular cognitive systems, some of which was already underway, particularly in linguistics and in vision research. In fact, we might more accurately represent the State of Things at that time in this way: Fodor's book drew some general morals about the nature and future of cognitive science from several of the research programs in cognitive science that he had been immersed in over the preceding dozen years or so. But these general morals themselves became instrumental in directing research in a variety of areas. Included here are developmental psychology—from the idea of a naïve physics, to folk biology, to the theory of mind module; cognitive neuroscience—involving the search for the neural underpinnings of Fodorean (what we might call *cognitiv*e) modularity theory as well as the development of a more fine-grained, smaller-scale notion of modularity; and evolutionary psychology.

The specific form of cognitive dissonance for the aficionado of Fodoriana is this: Fodor's argument in *Mind* does little more than spell out in more explicit detail the general argument underlying the second part of his two-fold modularity thesis, and as such, is directed not only at evolutionary psychology but at other extensions of the notion of modularity beyond what Fodor views as their proper reach. The book could, then, simply be read as a slap on the wrist for those who had not taken Fodor's First Law seriously enough. But what I find puzzling is that there is essentially *no* reference in *Mind* to *anything* that Fodor finds remotely plausible in cognitive science over the past 20 years, either in support of his First Law, or in response to objections to it. And this is puzzling because it is simply hard to believe that *nothing* in the literature in 20 years could speak in favour or against an adventuresome, speculative proposal about the nature of cognition and its study (cf. Jackendoff 2002). What's been going on? I think two distinct things.

First, Fodor's initial arguments, like his more recent ones regarding modularity, proceed largely *a priori*, i.e., they don't turn on empirical details. In *Mind*, there's the putative incompatibility between local computations and global effects of cognition (ch.2), and what Fodor calls *the input problem* (pp. 71ff.), neither of which turns on any empirical details. In *Modularity*, there was the charge that central processes were *Quinean* and *isotropic*, based largely on an analogy to claims about the nature of scientific confirmation, rather than an analysis of empirical work on any such processes, such as reasoning, problem-solving, or decision-making.

Second, large areas of developmental and cognitive psychology are now quite far down paths that follow one of the wrong turns that Fodor thinks that cognitive science has taken—the extension of modularity theory beyond its proper domains. As also intimated above, Fodor also doesn't care much for the neuroscientific turn of the '90s (remember the Decade of the Brain?), nor for many of the other shifts in the climate of cognitive science, including dynamic approaches to cognitive phenomena and the recent work on the embodiment of cognition (e.g., Clark 1997, 2001; Port and van Gelder 1995). Since Fodor also thinks little of other West Coast enthusiasms, such as cognitive linguistics, and has savaged philosophers, psychologists, and linguists for their attempts to understand concept acquisition and conceptual structure (Fodor 1998), there's not much left for him to draw on from the recent cognitive sciences.

As an aside, Fodor has my partial sympathy vis-à-vis neuroscience, insofar as it is typically *very* difficult to read off conclusions about large-scale cognitive structure from relatively small-scale findings in the neurosciences (despite the rabid enthusiasm for doing so within cognitive neuroscience). Even with the recent advancements in imaging technologies, particularly fMRI and PET, there is hardly ever a direct path linking larger-scale findings in cognitive neuroscience to claims about cognitive architecture. There has been a flurry of experimental results using such technology that purport to find the neural basis for a given cognitive ability, skill, or process. But as Dan Lloyd (2000) has recently pointed out, without a cross-study analysis of what sorts of abilities, skills, and processes tend to be associated with activity in which brain regions, and vice-versa, individual studies here show little, except perhaps the researcher's localistic biases in their research methodology (cf. also Uttal 2001).

I shall aim to do two things in the body of the paper before coming back to grapple with the larger question of "what happened?" The first will be to make some brief points about Fodor's two chief arguments for the sceptical part of his modularity thesis (sections 3 and 4). Following that, I shall identify some work in cognitive science that does seem relevant for assessing the conclusion of those arguments (section 5).

3. Argument 1: Local Computation, Global Cognition

To assess the claim that there is an inherent tension between the local nature of computational processes and the (occasional) global effects of cognition, we should begin with two questions:

1 Is computation local?
2 Are there global effects of cognition?

Is computation local? Fodor has returned, over the years, to the idea that computation is local, and in *Mind* he approaches this via the idea that the syntactic properties of representations are local,

> which is to say that they [syntactic properties] are constituted entirely by what parts a representation has and how these parts are arranged. You don't have to look 'outside' a sentence to see what its syntactic structure is, any more than you have to look outside a word to see how it is spelled. (p.20)

To say that computations are local is to say that the representations they operate over are "constituted entirely" by their parts and the arrangements of those parts. To deny this is to claim that there are some factors beyond a given representation, R, that in part "constitute" it.

There is a trivial reading of this claim about computation that should be put to one side, one that takes "constitution" to refer to or be treated as *physical* constitution. On this reading, claiming that computations are local would be only to claim that their representations have parts related in certain ways, and that those parts and those relations together physically constitute the representation. This claim has no modal dimension, and while it may be too flippant to characterize this as a completely trivial claim, it is one that few would get excited about. There is something stronger that I think Fodor means to imply in claiming that computation is local.

That something stronger—needed in fact to generate a *prima facie* conflict between computation and abduction—can be expressed in the dual language of determination and supervenience: it is that the "here and now" properties of representations provide a determining, subvenient base for the computational operations performed on them, at least *qua* computational processes. This is a substantial thesis in that it implies that the only relational properties relevant to computations involving R are those that are or are determined by the parts of R and their relations to one another. So construed, computation is *individualistic*.

This view of computation, popular as it is, is mistaken. I originally argued this, in effect, in developing a *wide computational* alternative to mainstream computationalism, which makes this assumption that computation is local, and hence individualistic (Wilson 1994, 1995:ch.3). There I had conceived of wide computationalism as a simple extension of the local view of computation to computational systems that extend beyond the boundary of the individual cognizer, but I now think that wide computationalism has some more radical implications for how we think about computation, and whether it is "local".

Wide computationalism is the view that at least some of the computations that individual cognizers perform extend beyond the boundary of those cognizers. Motivating wide computationalism is a conjunction of two ideas: first, that the notion of a computational *system* is basic, with particular computational relations and processes characterized in terms of it; and second, that (cognitive) computational systems and individual agents can stand in either a part-whole *or* a whole-part relation, just which depending on the details of the particular cognitive system being considered. In terms I have used elsewhere (Wilson 2001), computational systems can be realized either as *entity-bounded* or as *wide* systems, i.e., in systems that physically extend beyond the boundary of the individual cognizer.

Since I have discussed the second of these two motivating ideas in detail elsewhere, here I want to attend to the first idea—that of the primacy of the notion of a computational *system*—in the context of a discussion of some broader metaphysical questions. I don't think that the "local" view of computation falls out of Turing's work on computation (despite Fodor's own comments e.g., *Mind*, ch.1), but owes much of its force instead to broader metaphysical views. And I think that we need to move beneath the glosses on "local" as syntactic, non-semantic, formal, mechanical, physical, etc. that Fodor has provided over the years. Going all metaphysical at this point seems to me necessary to assess the idea that computation is local.

The idea that computational properties, as a type of mechanical or causally efficacious property, are or are determined by the intrinsic properties of individual representational tokens, together with the relations between such tokens, is a species of a view that I have dubbed *smallism* (Wilson 1999, 2004)**:** discrimination in favor of the small, and so against the not-so-small. Smallism is a sort of global metaphysics that has played an influential role in contemporary physicalism in the philosophy of mind and in general philosophy of science. Or, rather, it is a general attitude that lies in the background of a number of such global metaphysical views, such as theses of microstructural and Humean supervenience. David Lewis eloquently expresses the latter doctrine in the preface to volume II of his collected papers:

> all there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another ... we have local qualities: perfectly natural intrinsic properties which need

> nothing bigger than a point at which to be instantiated. For short: we have an arrangement of qualities. And that is all. (p.ix, 1986).

The question to be asked of all such views is what account they provide of relational properties. As Lewis's talk of an "arrangement of qualities" suggests, smallist views typically appeal to the relations that hold between very small things with intrinsic properties, and claim that all properties and relations are determined by these. But there is a systematic problem such views face. Suppose that A and B are related via R to form C. Then *perhaps* it is true that all of C's intrinsic properties are determined by the intrinsic properties of A, B, and R. But C's *relational* properties won't be so determined, but will be determined in part by things extrinsic to C. Smallist views are typically *aggregative* in this respect, and the problem with this sort of aggregative determination is that, if there are relational properties of the aggregation, they can't be accounted for solely in terms of the constitution of the aggregation.

The same holds true of the view that computation is local, and is one of the things that drives the systemic view of the nature of computation. Whether any given physical property of a representational token is a computational property depends on facts about the broader computational system in which it functions, including the nature of the code it uses and what are usually thought of as implementational details (e.g., compilation, configuration). Fodor has, over a long period of time, referred to "shape" as a paradigm of a computational property of mental representations. But particular computational systems are sensitive only to particular shapes, and so whether a *given* shape (i.e., an instantiated, determinate shape) functions computationally can vary from system to system, and so *qua* computational property is *not* an intrinsic property of the tokens that have it. This is true of computational systems at both the "highest" levels (e.g., word-processing systems) and the "lowest" levels (e.g., binary). True, something shaped as an "A", together with something shaped as "A→B" will lead, computationally, to something shaped as "B". But only in a computational system that includes modus ponens (or some such rule).

There is some affinity between the systemic view of computation and some rather extreme departures from standard views of computation, but we should be clear on both the affinities and the differences here. John Searle (1992) has claimed that syntax was an ascriber-relative property in deepening his argument against the computational theory of mind. Likewise, Steven Horst (1996) has defended a semiotic analysis of "symbol" in arguing that the ascription of computational status presupposes semiotic conventions. What these views share with the systemic view of computation is the denial that small, particulate bits of the physical world are themselves symbolic, syntactic, formal, or computational. Where these views differ is in seeing this as the basis for rejecting computationalism altogether, rather than modifying it to reflect the fact that larger bits of the physical world, structured so as to contain parts that correspond to the parts of formal systems, such as symbols and rules linking them, *are* computational in and of themselves. There really are things that compute, but those things are systems of entities, with the computational status of those entities being derivative from that of the systems in which they operate.

Are there global effects of cognition? In *Modularity*, Fodor argued that "central processes" were Quinean and isotropic—i.e., sensitive in degree and kind to global properties of cognitive systems—via an analogy to scientific confirmation. In *Mind* Fodor makes much the same claim without relying as explicitly on the analogy to confirmation. The common problem shared by both versions of Fodor's claim here is that it remains extremely unclear just what a "global effect" of cognition or a "global property" of a cognitive system is. Fodor says that "[S]implicity is, I think, a convincing example of a context-dependent property of mental representations to which cognitive processes are responsive" (*Mind*, p.25), and goes on (pp.33-37) to discuss *conservatism* (i.e., not changing your beliefs without reason) as another example.

In both cases, Fodor begins by pointing to the *normative* roles that each of these notions plays in accounts of rationality, appealing to simplicity and conservativism as "part of rationality" and "constitutive of rationality", respectively. He then moves on to the claim that as cognizers we can and do make judgments of simplicity, and in fact are conservative in belief change, and asks how this is possible, given the computational nature of thought. Let us take these points one at a time.

First, the normative point seems simply irrelevant to the question of whether there are global effects of cognition. That there are norms governing rationality that appeal to "global properties" of cognitive processes, such as simplicity or conservativism, tells us nothing about whether cognition abides by those norms. Notoriously, standard accounts of rationality are cast in terms of norms, such as various forms of optimization or maximization, that cognizers like us with limited resources can at best approximate. So the burden of Fodor's argument falls on his claim that we are actually sensitive to properties such as simplicity and conservativism in adjusting our cognitive sets.

But this point about what properties of propositions and claims we are actually sensitive to is, I think, more difficult to connect to the claim that there are "global effects" of cognition than Fodor supposes. We make judgments about all sorts of things, and use a variety of criteria to adjust the contents of our minds. It is extremely unclear how these facts amount to cognition being "global". Consider simplicity, a property we attribute to certain propositions, theories, or claims. The question of *how* we do that—a question to which the computational theory of mind is a general answer—is quite distinct. Likewise, conservativism is a tendency that (suppose) our cognition exhibits. That is a property of our cognitive set as a whole, or significant chunks of it. That it is, in some sense, global implies nothing about how that tendency is manifested or how conservative change takes place. In short, Fodor's "global effects of cognition" are *products* of cognitive processes, not features of cognitive processes themselves. Moreover, unless one supposes that processes that generate an effect must share the properties that those effects have—in this case, being global—there seems no basis for making claims about the processes based on claims about features of the products of those processes.

4. Argument 2: The Input Problem

Fodor poses the input problem as a problem for a proponent of the view that the entire mind is modular in nature, arguing that "[M]echanisms that operate as modules *presuppose* mechanisms that don't" (p.71). Although the argument is run as a dilemma by Fodor on the question of whether inputs to modules are determined to have the properties relevant to their status as inputs by one or two distinct mechanisms, it can be expressed succinctly as follows. There must be some process that determines whether a given representation has the properties sufficient for it to be an input to any given modular process, M. But that input-determining process must be less modular than M, for it is a mechanism, in effect, for selecting or creating input representations for M from representations more generally. Supposing that there are distinct such input determining mechanisms for each module only defers the problem, since the same is true of whichever mechanism provides *its* inputs. Either way, there must be parts of the mind that are less modular than any given module in order for there to be identifiable inputs to that module. Hence, not all of the mind can be modular.

This argument in fact entails only that not all of the mind can be *equally* modular; it is compatible with all of the mind being modular, and differing by degree in the level of modularity that its various parts have. Yet there is something strange about the argument, so modified, since what it strictly implies is that each cognitive process, however modular it is, must be "fed" by a less modular process, and so the least modular processes are those that process the most basic inputs, i.e., perceptual input systems (or even transducers, in Fodor's [1983] terms). But these are our (and Fodor's) paradigmatic modular systems, so the conclusion that they are less modular than "downstream" modules is *prima facie* surprising.

The problem, I think, lies in a dual failure on Fodor's part: his exclusive focus on inputs to the exclusion of outputs, and the linear, more-to-less (or general-to-specific) construal of the temporal dimension to cognitive processing. Fodor focuses on the nature of a mechanism's inputs since that is directly related to its *domain-specificity*: the narrower the range of inputs to which it responds, the more domain-specific, and hence modular, it is. But a highly modular mechanism can still be complicated in terms of the range of outputs it generates, and this means that just a few domain-specific mechanisms can feed a large number of modules. Thus, the chain of processing is not necessarily from domain-general to domain-specific, with a pattern of reticulation corresponding to the temporal sequence of cognitive operations. Rather, this chain might form a mosaic of branching and reticulating trees, where those that branch can be as domain-specific and hence as modular as those that reticulate.

5. How Full is the Glass?: Some Empirical Considerations

So are we likely to have a computational, modular theory of intuitively "central processes", such as decision-making or reasoning, or does Fodor's First Law really hold? It is not just the specific arguments that Fodor offers that I have criticized in sections 3 and 4 that do not settle this issue, but *a priori* arguments more generally. It is a broadly empirical issue. So what sorts of empirical considerations do I think are most relevant to settling it? In this section, I mention three.

These considerations are not meant to respond to Fodor's arguments for a measure of scepticism about "massive modularity", since I have indicated already that those arguments are not empirical in nature, and in any case, are not very good arguments (cf. Sperber 1994, 2001). Rather, I want to point to work on cognition whose consideration should play a role in determining whether to share Fodor's scepticism about any sort of generalized modularity thesis, whether it be the "massive modularity thesis" of evolutionary psychology or the "core domains" view within developmental psychology (Hirschfield and Gelman 1994; Sperber, Premack and Premack 1995). That is, once we get beyond *a priori* arguments for or against the modularity of "central processes", these are areas of cognitive science whose details surely are relevant to assessing claims about whether a modular approach to understanding all of the mind is hopelessly confused.

Developmental Neuroscience. For the most part, developmental neuroscience has been ignored by both Fodor and by evolutionary psychologists. I think this is for two reasons. First, both have lent heavily on the thesis that psychology is autonomous from neuroscience and on the corresponding distinction between higher-level laws and lower-level mechanisms. Second, both have conceptualized modules as innate, hard-wired mechanisms of the mind. In so doing they have assumed that any developmental detail would simply inform us about the triggering conditions for the activation of innately-structured modules, or lay out how modules unfold over time.

Neither of these assumptions remains tenable in light of the complexities of brain development revealed within developmental neuroscience. In general terms, the human brain quadruples its size after birth, with this increase being distributed unevenly throughout the brain.

The neocortex, that part of the cortex supposedly shared by all mammals, increases most dramatically, and in fact is disproportionately larger in humans than in non-human mammals. In terms of the specific developmental trajectories that the cortex undergoes, it manifests both patterns of equipotentiality and more, special-purpose, dedicated neural pathways. For example, the visual cortex can process auditory input projected to the thalamus, but there is already widespread differentiation in the cortex before there is innervation from extracortical structures, suggesting a built-in specialization of cortical areas from the start (Levitt, Barbe, and Eagleson 1997; Finlay and Niederer 1999).

Such facts give pause to the sort of linkage made by Fodor and the evolutionary psychologists he criticizes between modularity and nativism. Since the structure of the brain itself emerges over time in a variety of experiential-dependent, and experiential-neutral ways (see also Buller and Hardcastle 2000, Quartz 2003), attention to the actual patterns of neural development and neural plasticity would seem critical for assessing any particular claims about the modular structure of cognition. Were the post-natal development of the human brain minor, or the effects of experience on its development relatively uniform, then both assumptions could be defended as methodological simplifications. Our minds might have been like the software programs installed on many desktop computers that simply need a few parameters to be set before they spring into action. Instead, however, it is as if whatever programming is already built in not only alters the very kind of computer we end up with, but that the program itself is significantly adjusted in light of user interaction, and this varies greatly across different parts of the computer. If that is more like our situation, as developmental neuroscience itself suggests, then the developmental details simply cannot be bracketed in considering questions of modularity.

Kludgy Cognitive Modeling. Fodor himself sees the frame problem as central to cognitive science, and as a beacon for any murky claims about the reasoning, problem-solving, or inferential capacities of cognitive models, especially those within classical artificial intelligence. Yet the frame problem is a problem equally for (a) human beings, and (b) actual computational systems that in fact work pretty well. The fact that both of these manage to avoid the frame problem, or to solve it, should make us question the status that Fodor ascribes to that problem. There is an analogy here to the problem of induction: it is not that there is no real problem with induction, but it would be a mistake to conclude from that problem that people (or machines, for that matter), can't perform inductions. Clearly, they do. In both cases, we can construe the problem as setting some sort of normative ideal, the path to which seems inherently problematic.

But rather than keep either problem exclusively in that light, one might attend instead to how actual systems manage to kludge well enough to solve the corresponding problem well-enough enough of the time. Although the notion of a kludge was introduced in artificial intelligence with more than a hint of being something to avoid, in fact kludges are simply the natural product of heuristically-driven, pragmatic solutions to cognitive (or other) problems. In problem solving and decision-making, Herb Simon's (1969) ideas of bounded rationality and satisficing, and the theories they have generated, are widely accepted, and they were introduced in opposition to theories of cognitive performance, such as rational choice theory, tied to unachievable normative goals (see also Gigerenzer and Selten 2001). Optimality does not drop out of such views here, but is reconceptualized in terms of sets of constraints and bounds on cognition. This is no different from how optimization is reconceptualized within current adaptationist paradigms in evolutionary biology (Orzack and Sober 2001).

So perhaps the mind is primarily kludgy, and areas within cognitive science to explore more fully in evaluating Fodor's First Law and the pessimism it expresses are those that take kludges more seriously. Included here would be work on embodied cognition, drawing on constraints derived from the fact that minds operate in or via bodies (Clark 1997, Grush 2003); work within cognitive neuroscience that builds on constraints derived from neural processing (Glimcher 2003), and work in computational intelligence on induction and learning, especially that building on computational learning theory (Valiant 1984, Dietterich 1990).

Induction, Analogy, and Inference. Those outside of cognitive science encountering Fodor's claim that there is an inherent tension between computation and abduction may be surprised to know that there is in fact a considerable body of empirical and theoretical work on induction, analogy and inference conducted within a computational framework. It *may* be that Fodor's pessimism about the history of artificial intelligence is justified, but without discussion of any real examples of modeling and engineering work, this is difficult to see. Turning to some examples of work in general areas that I think has made some progress and provided some insight into "central" cognitive processes, particularly inductive or abductive processing, consider two.

First, there is Paul Thagard's work (e.g., 2000) on *coherence* as constraint satisfaction, and its application to a range of problems that human beings solve most likely non-deductively. Thagard acknowledges the "in principle" limits that exist to the computational solution of NP-hard problems, i.e., problems that are intrinsically difficult for a non-deterministic Turing machine to solve in polynomial time. But, more importantly, he also indicates a range of approximative techniques, such as harmony maximization in connectionist networks (e.g., in ECHO) and greedy local search in classic cognitive architectures, that allow actual systems to produce viable and plausible solutions to such problems. Of course, such strategies are not optimal in the sense that they are not guaranteed to find the best possible solution; but there is also no evidence that I know of that *we* do so, and much evidence to suggest that we don't.

Second, some of the most interesting work currently being done on inference occurs at the intersection of classical and connectionist approaches, particularly work on Bayesian networks and machine learning (see Jordan and Russell 1999, Pearl 2000). The general idea is to integrate causal and probabilistic considerations into existing frameworks to produce computational systems that are more flexible and context-sensitive. For all that, they are no less mechanistic in how they operate, and solve, in specific contexts, abductive, "global" problems through standard computational means.

6. Relieving Some Cognitive Dissonance: Jerry Fodor as …

So what has happened to the Jerry Fodor who put the Quinean view of philosophy as continuous with the natural sciences into practice in thinking about the mind? It's a platitude of folk psychology that people change, and one we might adopt here to relieve a little cognitive dissonance. Fodor has changed. (I feel better already.) But into whom?

Jerry Fodor as Colin McGinn? Fiona Cowie (1999) has argued that there is a mysterian strand to Fodor's views of nativism about concepts. There's something that no theory of concepts can tell us—how concepts are acquired—and this limitation is principled, in much the way that there's just something about consciousness that we'll never manage to get our minds around (McGinn 1991). Maybe one way to understand Fodor is to view him as holding that much of the mind is like that.

Jerry Fodor as Hubert Dreyfus? Although Dreyfus's *What Computers Can't Do* (1972) might be thought as an indictment of the computationalism that forms the core of Fodor's views of the mind, through-a-glass-darkly its basic message is not that different from that of *Mind*. Computational approaches to understanding the mind are quite limited. True, Dreyfus is up to his armpits in *Dasein*, something that Fodor doesn't even like dipping his big toe into, and turns out to be something of a connectionist groupie. But just as Dreyfus reinforced his basic views of the limitations of a computer-oriented view of cognition 20 years later in *What Computers Still Can't Do* (1992), Fodor has done the same with his view of the limitations of computational psychology in *Mind*.

Jerry Fodor as Granny? Fodor has always been fond of his Granny. Granny, the defender of folk wisdom about the mind, has rocked in her chair these last 40 years, and Fodor has dedicated no inconsiderable amount of his considerable energy to showing that his views of the mind give Granny only reason to smile. But, as Janet Leigh found out at the Bates Motel, sometimes filial fondness comes with its own price. Has Fodor *become* Granny, defender of the status quo and content to rock and say "I told you so"? (cf. Dennett 1991).

I close with what we might see as Fodor's own diagnosis (1991:280), offered in reply to Dennett:

> My view about the psychology of central processes is not that it's impossible in principle, and of course it's not that 'scientists should be [prohibited] from attempting empirical explorations …'. My view is this: there are some problems you can't solve because key ideas are missing. Blustering doesn't help, throwing money at the

> problems doesn't help, arguments of the form 'some theory must work, this is some theory, therefore this theory must work' don't help (although that's often quite a good form of argument); nothing helps until somebody gets some key ideas.

So, wanted: a few good ideas. The only question is where we might find these.

References

Buller, D. and V.G. Hardcastle, "Evolutionary Psychology, Meet the Developing Brain: Combating Promiscuous Modularity", *Brain and Mind* 1:307-325.

Clark, A., 1997, *Being There: Putting Brain, Body, and World Together Again*. Cambridge, MA: MIT Press.

-----2001, "Reasons, Robots and the Extended Mind", *Mind and Language*, 16:121-145.

Cowie, F., 1999, *What's Within? Nativism Reconsidered*. New York: Oxford University Press.

Dennett, D.C., 1991, "Granny's Campaign for Safe Science", in B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and his Critics*. Cambridge, MA: Blackwell.

Dietterich, T.G., 1990, "Machine Learning", *Annual Review of Computer Science* 4:255-306.

Dreyfus, H.L., 1972, *What Computers Can't Do: A Critique of Artificial Reason*. New York: Harper and Row.

-----1992, *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.

Finlay, B. and J.K. Niederer, 1999, "Neural Development", in R.A. Wilson and F.C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. MIT Press, pp.594-596.

J.A. Fodor, 1975, *The Language of Thought*. Cambridge, MA: Harvard University Press.

-----1981, *Representations*. Sussex, England: Harvester Press.

-----1983, *The Modularity of Mind*. Cambridge, MA: MIT Press.

-----1987, *Psychosemantics*. Cambridge, MA: MIT Press.

-----1990, *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

-----1991, "Replies", in B. Loewer and G. Rey (eds.), *Meaning in Mind: Fodor and his Critics*. Cambridge, MA: Blackwell.

-----1994, *The Elm and the Expert*. Cambridge, MA: MIT Press.

-----1998a, *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*. Cambridge, MA: MIT Press.

-----1998b, *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.

-----2000, *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.

Gigerenzer, G. and R. Selten (eds.), 2001, *Bounded Rationality: The Adaptive Toolbox*. Cambridge, MA: MIT Press.

Glimcher, P.W., 2003, *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. Cambridge, MA: MIT Press.

Grush , R., 2003, "In Defense of Some 'Cartesian' Assumptions Concerning the Brain and its Operation", *Biology and Philosophy* 18:53-93.

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

Horst, S., 1996, *Symbols, Computation, and Intentionality: A Critique of the Computational Theory of Mind*. Berkeley, CA: University of California Press.

Jackendoff, R., 2002, Review of *The mind doesn't work that way*, *Language* 78:164-170.

Jordan, M. and S. Russell, "Computational Intelligence", in R. A. Wilson and F.C. Keil (eds.), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MA: MIT Press, pp. lxxiii-xc.

Levitt, P.M., F. Barbe, and K.L. Eagleson, 1997, "Patterning and Specification of the Cerebral Cortex", *Annual Review of Neuroscience* 17:109-132.

Lewis, D.K., 1986, *Philosophical Papers, Volume II*. Oxford: Oxford University Press.

McGinn, C., 1991, *The Problem of Consciousness*. Oxford: Blackwell.

Okasha, S., 2003, "Fodor on Cognition, Modularity, and Adaptationism", *Philosophy of Science* 70:68-88.

Orzack, S., and E. Sober (eds.), 2001, *Adaptationism and Optimality*. New York: Cambridge University Press.

Pearl, Judea, 2000, *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Port, R. and T. van Gelder (eds.), 1995, *Mind as Motion*. Cambridge, MA: MIT Press.

Quartz, S., 2003, "Innateness and the Brain", *Biology and Philosophy* 18:13-40.

Searle, J.R., 1992, *Rediscovery of the Mind*. Cambridge, MA: MIT Press.

Simon, H.A., 1969, *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

Sperber, D., 1994, “The Modularity of Thought and the Epidemiology of Representations”, in L.A. Hirschfeld and S.A. Gelman (eds.), *Mapping the Mind: Domain Specificity in Cognition and Culture*. New York: Cambridge University Press, pp.39-67.

-----2001, “In Defense of Massive Modularity”, in E. Dupoux (ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*. Cambridge, MA: MIT Press, pp.47-57.

Sperber, D., D. Premack, and A. Premack (eds.), 1995, *Causal Cognition: A Multidisciplinary Debate*. New York: Oxford University Press.

Thagard, P., 2000, *Coherence in Thought and Action*. Cambridge, MA: MIT Press.

Uttal, W.R., 2001, *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*. Cambridge, MA: MIT Press.

Valiant, L.,1984, “A Theory of the Learnable”, *Communications of the ACM* 27:1134-1142.

Wilson, R.A., 1994, "Wide Computationalism", *Mind* 103:351-372.

-----*Cartesian Psychology and Physical Minds: Individualism and the Sciences of the Mind*. New York: Cambridge University Press.

-----1999, "The Individual in Biology and Psychology", in V.G. Hardcastle (ed.), *Where Biology Meets Psychology: Philosophical Essays*, MIT Press, Cambridge, MA, 1999, pp.357-74.

-----2001, "Two Views of Realization", *Philosophical Studies* 104:1-30.

-----2004, *Boundaries of the Mind: The Individual in the Fragile Science: Cognition*. New York: Cambridge University Press.