Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches[1]

Gualtiero Piccinini, University of Missouri – St. Louis

Carl Craver, Washington University in St. Louis

1/27/2011

This is a preprint of an article whose final and definitive form will be published in *Synthese*; *Synthese* is available online at: http://www.springer.com/philosophy/epistemology+and+philosophy+of+science/journal/11229

Abstract. We sketch a framework for building a unified science of cognition. This unification is achieved by showing how functional analyses of cognitive capacities can be integrated with the multilevel mechanistic explanations of neural systems. The core idea is that functional analyses are *sketches of mechanisms*, in which some structural aspects of a mechanistic explanation are omitted. Once the missing aspects are filled in, a functional analysis turns into a full-blown mechanistic explanation. By this process, functional analyses are seamlessly integrated with multilevel mechanistic explanations.

---

[1] This paper was refereed in two rounds. John Symons graciously arranged for a round of blind refereeing; thanks are due to him and the anonymous referees. In addition, thanks to those who refereed the paper non-blindly: Ken Aizawa, Robert Cummins, and Dan Weiskopf. Many thanks to audiences at the 2010 APA Pacific Division, Saint Louis University, and University of Missouri – Columbia. This material is based upon work supported by the National Science Foundation under Grant No. SES-0924527 to Gualtiero Piccinini.

1. Integrating Psychology and Neuroscience via Multi-Level Mechanistic Explanation

When psychologists explain behavior, the explanations typically make reference to causes that precede the behavior and make a difference to whether and how it occurs. For instance, they explain that Anna ducked because she saw a looming ball. By contrast, when psychologists explain psychological capacities such as stereopsis or working memory, they typically do so by showing that these complex capacities are made up of more basic capacities organized together. In this paper, we focus exclusively on the latter sort of explanation, which is usually referred to as *functional analysis*. We argue that such decompositional, constitutive explanations gain their explanatory force by describing mechanisms (even approximately and with idealization) and, conversely, that they lack explanatory force to the extent that they fail to describe mechanisms.

In arguing for this point, we sketch a framework for building a unified science of cognition. This unification is achieved by showing how functional analyses of cognitive capacities can be and in some cases have been integrated with the multilevel mechanistic explanations of neural systems. The core idea is that functional analyses are *sketches of mechanisms*, in which some structural aspects of a mechanistic explanation are omitted. Once the missing aspects are filled in, a functional analysis turns into a full-blown mechanistic explanation. By this process, functional analyses are seamlessly integrated with multilevel mechanistic explanations.

The conclusion that functional analyses are mechanism sketches leads to a simple argument that psychological explanation is mechanistic. It is generally assumed that psychological explanation is functional—that it proceeds via the functional analysis of cognitive capacities (Fodor 1968a, b; Dennett 1978, Chaps. 5 and 7; Cummins 1983, 2000; Block and Segal 1998). If psychological explanation is functional and functional analyses are mechanism sketches, then psychological explanations are mechanism sketches. Mechanism sketches are elliptical or incomplete mechanistic explanations. Therefore, psychological explanations are mechanistic.

This further conclusion conflicts with the way psychological explanation is traditionally understood. The received view is that functional analysis is autonomous and thus distinct from mechanistic explanation (Fodor 1965, 1968, 1974; Cummins 1983, 2000).[2] If psychological explanation is functional, neuroscientific explanation is mechanistic (Craver 2007), and functional analysis is distinct and autonomous from mechanistic explanation, then psychological explanation is distinct and autonomous from neuroscientific explanation. As we argue, though, functional and mechanistic explanations are not distinct and autonomous from one another precisely because functional analysis, properly constrained, is a kind of mechanistic explanation—an *elliptical* mechanistic explanation. Thus, a correct understanding of functional analysis undermines the influential claim that explanation in psychology is distinct and autonomous from explanation in neuroscience.

[2] E.g.: "[V]is-à-vis explanations of behavior, neurological theories specify mechanisms and psychological theories do not" (Fodor 1965, 177)

Our rejection of the autonomy of psychological explanation should not be confused with a rejection of multiple realizability or an endorsement of reductionism. Several authors argue that functional or psychological properties are not multiply realizable—or that if functional or psychological properties are multiply realizable, then they are not natural kinds (Bechtel 2009; Bechtel and Mundale 1999; Bickle 2003, 2006; P.M. Churchland 2007; Couch 2005; Kim 1992; Polger 2004, 2009; Shagrir 2008; Shapiro 2000, 2004). Many of these authors conclude that psychological explanations either reduce to or ought to be replaced by neuroscientific explanations (Kim 1992, P.M. Churchland 1989, 2007, P.S. Churchland 1986, Bickle 2003). In response, others have defended the multiple realizability of functional or psychological properties, usually in conjunction with a defense of the autonomy of psychology (Aizawa and Gillett 2009, 2011, forthcoming; Block 1997; Figdor 2010; Fodor 1997; Gold and Stoljar 1999).

Our rejection of the autonomy thesis is in many ways orthogonal to traditional debates about whether psychological properties or functions are multiply realizable. Even if functional properties are multiply realizable, functional analysis is still a kind of mechanistic explanation; a fortiori, functional analysis is not autonomous from mechanistic explanation, and psychological explanation is not autonomous from neuroscientific explanation.

In fact, there is a kind of multiple realizability—multiple functional decompositions of the same capacity—that militates *against* the autonomy of psychology. Autonomist psychology—the search for functional analysis without direct constraints from neural structures—usually goes hand in hand with the assumption that each psychological capacity has a unique functional decomposition (which in turn may have multiple realizers). But there is evidence that the same psychological capacity is fulfilled at different times by entirely different neural structures, or different configurations of neural structures, even within the same organism (Edelman and Gally 2001, Friston and Price 2003, Noppeney et al. 2004, Figdor 2010). Plausibly each such configuration of neural structures corresponds to a somewhat different functional decomposition. So several functional decompositions may all be correct across different species, different members of the same species, and even different time-slices of an individual organism.

Yet the typical outcome of autonomist psychology is a single functional analysis of a given capacity. Even assuming for the sake of the argument that autonomist psychology stumbles on one among the correct functional analyses, autonomist psychology is bound to miss the other functional analyses that are also correct. The way around this problem is to let functional analysis be constrained by neural structures—that is, to abandon autonomist psychology in favor of integrating psychology and neuroscience. Thus, the multiplicity of functional decompositions—a kind of multiple realizability—does not support autonomy but rather undermines it. This being said, we will set multiple realizability aside.

Our argument against the autonomy thesis is *not* an argument for reductionism, either as it has been classically conceived (as the derivation of one theory from another) or as it is now commonly conceived (as the idea that lower-level mechanisms are explanatorily privileged).[3] Instead, our argument leads to a new understanding of how psychology and neuroscience should be integrated—explanatory unification will be achieved through the integration of findings from different areas of neuroscience and psychology into a description of multilevel mechanisms.

The most immediate consequence of the present integrationist program is that theorists of cognition ought to learn how nervous systems work and use that information to constrain their investigations. Of course, many theorists already do this. Indeed, recent decades have witnessed an increasing integration of psychology and neuroscience roughly along the lines advocated here. Many psychologists have been moving away from less mechanistically constrained models towards models that take more and more neuroscientific constraints into account (e.g., Gazzaniga 2009, Kalat 2008, Kosslyn 2006, O'Reilly and Munakata 2000, Posner 2004). Our goal is to express one rationale for that trend: there is no kind of constitutive explanation of psychological phenomena that is distinct from mechanistic explanation, properly conceived.

There are still psychologists who pursue explanations of cognition without concerning themselves with how nervous systems work and philosophers who question whether explanations that incorporate neuroscientific evidence (such as the localization of cognitive functions in the brain) are any better than explanations that ignore neuroscientific evidence. We disagree with each. Insofar as psychologists pursue

---

[3] We do endorse reductionism in the sense that every concrete thing is made out of physical components and the organized activities of a system's components explain the activities of the whole. Setting aside dualism and spooky versions of emergentism, we take these theses to be uncontroversial.

constitutive explanations, they ought to acknowledge that psychological explanations describe aspects of the same multilevel neural mechanism that neuroscientists study. Thus, psychologists ought to let knowledge of neural mechanisms constrain their hypotheses just like neuroscientists ought to let knowledge of psychological functions constrain theirs.

In the next section, we outline the received view that functional analysis is distinct and autonomous from mechanistic explanation. After that, Section 3 sketches the basic elements of mechanistic explanation, emphasizing that functional properties are an integral part of mechanisms. Sections 4-6 discuss the three main types of functional analysis, arguing that each is a sketch of a mechanism. Section 7 rounds up our argument by going over an example of how identifying a functional kind (neurotransmitter) within a system requires fitting it into a mechanism.

2. The Received View: Functional Analysis as Distinct and Autonomous from Mechanistic Explanation

There is consensus that psychological capacities are explained functionally, that is, by means of what is often called *functional analysis*. Functional analysis is the analysis of a capacity in terms of the functional properties of a system and their organization. Three main types of functional analysis may be distinguished, depending on which functional properties they invoke. One type is *task analysis*: the decomposition of a capacity into subcapacities and their organization (Cummins 1975, 1983, 2000; cf. also Fodor 1968b and Dennett 1978). A second type is *functional analysis by internal states*: an account of how a capacity is produced in terms of a set of internal states and their mutual interaction (Fodor 1965, 1968, Stich 1983; cf. also Putnam 1960, 1967a, b). A third type is *boxology*: the decomposition of a system into a set of functionally individuated components (black boxes), the processes they go through, and their organization (Fodor 1965, 1968).

We focus on these three types of functional analysis because they have been articulated in the greatest detail and defended most prominently in the literature on psychological explanation. We address each kind of functional analysis and argue that it amounts to a mechanism sketch. But our argument is not limited to these three kinds of functional analysis. If there are yet other types of functional analysis, our argument can be extended to cover them. For we argue that a complete constitutive explanation of a phenomenon in terms of functional properties requires identifying the structures that possess those functional properties—that is, it requires fitting the functional properties within a mechanism. Thus, while we focus on the above-mentioned types of functional analysis, our conclusion applies to all types of functional analysis.

By the same token, our argument applies to any combination of different types of functional analysis. For instance, a capacity of a system may be explained in terms of a system of subcapacities (task analysis) manifested by a set of appropriately organized black boxes (boxology) when the black boxes are in suitable internal states (functional analysis by internal states). Our argument entails that such explanations are not adequate and complete until the black boxes are identified with concrete structures and the internal states and subcapacities are identified with states and capacities of the structures that make up the system.

The received view of the relationship between functional analysis and mechanistic explanation may be summarized as follows:

> Distinctness: Functional analysis and mechanistic explanation are distinct kinds of explanation.
>
> Autonomy: Functional analysis and mechanistic explanation are autonomous from one another.

One way to see that proponents of the received view endorse distinctness is that they often claim that a complete explanation of a capacity includes both a functional analysis and a matching mechanistic explanation.[4] This presupposes that functional analyses are distinct from mechanistic explanation.

Distinctness is defended in slightly different ways depending on which form of functional analysis is at issue. With respect to task analysis, distinctness has been defended as follows. Unlike mechanistic explanation, which attributes subcapacities to the *components* of a mechanism, task analysis need not attribute subcapacities to the

---

[4] E.g.:

> Explanation in psychology consists of a functional analysis and a mechanistic analysis: a phase one theory and a determination of which model of the theory the nervous system of the organism represents (Fodor 1965, 177).
>
> Functional and mechanistic explanations must be matched to have a complete explanation of a capacity (Cummins 1983, 31).

components of the system because the subcapacities may all belong to the whole system.[5]

While talking about components simpliciter may be enough to distinguish between task analysis and mechanistic explanation, it is insufficient to distinguish between functional analysis in general and mechanistic explanation. This is because some analyses—specifically, functional analyses that appeal to black boxes—do appeal to components, although such components are supposed to be individuated purely functionally, by what they do. To avoid ambiguity, we use the term *functional components* for components as individuated by their functional properties and *structural components* for components as individuated by their structural properties.

Paradigmatic structural properties include the size, shape, location, and orientation of extended objects. Anatomists tend to study structures in this sense, as do x-ray crystallographers. Paradigmatic functional properties include being a neurotransmitter, encoding an episodic memory, or generating shape from shading. Physiologists tend to study function. Nothing in our argument turns on there being a metaphysically fundamental divide between functional and structural properties; indeed, if we are

---

[5] Cf. Cummins:

> Form-function correlation is certainly absent in many cases … and it is therefore important to keep functional analysis and componential analysis [i.e., mechanistic explanation] conceptually distinct. Componential analysis of computers, and probably brains, will typically yield components with capacities that do not figure in the analysis of capacities of the whole system (Cummins 1983, 2000, 125).

right, one cannot characterize functions without committing oneself to structures and vice versa.

Those who think of functional analysis in terms of internal states and boxology defend distinctness on the grounds that internal states or black boxes are individuated solely by their functional relations with each other as well as with inputs and outputs, and not by their structural properties. While mechanistic explanation appeals to the structural features of the components of the mechanism, functional analysis allegedly does not.[6]

Distinctness is a necessary condition for autonomy: if functional analysis is a kind of mechanistic explanation, as we argue, then functional analysis cannot be *autonomous from* mechanistic explanation. But distinctness is not *sufficient* for autonomy: two explanations may be distinct from each other and yet mutually dependent. Nevertheless, those who endorse distinctness typically endorse autonomy as well—in fact, we suspect that defending autonomy is the primary motivation for endorsing distinctness.[7]

---

[6] Cf. Fodor:

> If I speak of a device as a "camshaft," I am implicitly identifying it by reference to its physical structure, and so I am committed to the view that it exhibits a characteristic and specifiable decomposition into physical parts. But if I speak of the device as a "valve lifter," I am identifying it by reference to its function and I therefore undertake no such commitment (Fodor 1968, 113).

[7] E.g.:

> The conventional wisdom in the philosophy of mind is that psychological states are functional and the laws and theories that figure in psychological explanations are autonomous (Fodor 1997, 149).

What is autonomy (in the relevant sense)? There are many kinds of autonomy. One scientific enterprise may be called *autonomous* from another if the former can choose (i) which phenomena to explain, (ii) which observational and experimental techniques to use, (iii) which vocabulary to adopt, and (iv) the precise way in which evidence from the other field constraints its explanations. The term "autonomy" is sometimes used for one or more of the above (e.g., Aizawa and Gillett forthcoming argue that psychology is autonomous from neuroscience in sense (iv)). We have no issue with these forms of autonomy. Psychology may well be autonomous from neuroscience in these four ways.

In another sense, a scientific explanation may be said to be autonomous from another just in case the former refers to properties that are distinct from and irreducible to the properties referred to by the latter. This form of autonomy is sometimes claimed to obtain between psychology and neuroscience. For instance, psychological properties are sometimes claimed to be distinct from and irreducible to the neural properties that realize them.

This latter form of autonomy thesis suggests that properties are stacked into levels of being. It is not clear how levels of being can be distinct from one another without being ontologically redundant (Kim 1992, Heil 2003). Doing justice to this topic would take us

---

> Why … should not the kind predicates of the special sciences cross-classify the physical natural kinds? (Fodor 1975, 25; see also Fodor 1997, 161-2)
>
> We could be made of Swiss cheese and it wouldn't matter (Putnam 1975, 291).

It is worth noting that Fodor's writings on psychological explanation from the 1960s were less sanguine about autonomy than his later writings, although he was already defending distinctness (cf. Aizawa and Gillett forthcoming).

beyond the scope of this paper and is orthogonal to our topic. When we talk about mechanistic levels and levels of mechanistic explanation, we are officially neutral on whether mechanistic levels correspond to levels of being that are ontologically autonomous from one another.

In yet another sense, autonomy may be said to obtain between either laws or theories when they are irreducible to one another (cf. Fodor 1997, p. 149; Block 1997), regardless of whether such laws or theories refer to ontologically distinct levels of being. As we have pointed out before, we are not defending reductionism. Nevertheless, we reject this kind of autonomy. The dichotomy between reduction and autonomy is a false one. Neither psychology nor neuroscience discovers the kind of law or theory for which talk of reduction makes the most sense (cf. Cummins 2000). What they discover, we argue, are aspects of mechanisms to be combined in full-blown multilevel mechanistic explanations. Psychological explanations are not distinct from neuroscientific ones; each describes aspects of the same multilevel mechanisms. Therefore, we reject autonomy as irreducibility of laws or theories in favor not of reduction but of explanatory integration.

A final kind of autonomy thesis maintains that two explanations are autonomous just in case there are no *direct constraints* between them. Specifically, some authors maintain that the functional analysis and the mechanistic explanation of one and the same phenomenon put no direct constraints on each other.[8] While proponents of this kind of

---

[8] E.g.:

autonomy are not very explicit in what they mean by "direct constraints," the following seems to capture their usage: a functional analysis directly constrains a mechanistic explanation if and only if the functional properties described by a functional analysis restrict the range of structural components and component organizations that might exhibit those capacities; a mechanistic explanation directly constrains a functional analysis if and only if the structural components and component organization described by the mechanistic explanation restrict the range of functional properties exhibited by those components thus organized.

Of course, every participant in this debate agrees on one important (if obvious) *indirect* constraint: the mechanism postulated by a true mechanistic explanation must *realize* the functional system postulated by a true functional analysis.[9] Aside from that,

---

> Phase one explanations [i.e., functional analyses by internal states] purport to account for behaviour in terms of internal states, but they give no information whatever about the mechanisms underlying these states (Fodor 1965, 173).
>
> [F]unctional analysis puts very indirect constraints on componential analysis (Cummins 1983, 29; 2000, 126).

While these specific statements by Cummins and Fodor entail the autonomy of mechanistic explanation from functional analysis rather than the converse, the rest of what they write makes clear that they also maintain that functional analysis is autonomous from mechanistic explanation. For an even stronger formulation of the "no constraint principle" that is pervasive in the literature on functional analysis, see Aizawa and Gillett forthcoming.

[9] Cf. Cummins:

> Ultimately, of course, a complete theory for a capacity must exhibit the details of the target capacity's realization in the system (or system type) that has it. Functional analysis of a capacity must eventually terminate in dispositions whose realizations are explicable via analysis of the target system. Failing this, we have no reason to suppose we have analyzed the capacity as it is realized in that system (Cummins 1983, 31; 2000, 126).

autonomists suggest that the two explanatory enterprises proceed independently of one another.[10] As this "no-direct-constraints" kind of autonomy is generally interpreted, it entails, or at least strongly suggests, that those who are engaged in functional analysis need not know or pay attention to what mechanisms are present in the system; by the same token, those who are engaged in mechanistic explanation need not know or pay attention to how a system is functionally analyzed. Thus, according to this kind of autonomy, psychologists and neuroscientists need not pay attention to what the other group is doing—except that in the end, of course, their explanations ought to match.

The assumption of autonomy as lack of direct constraints is appealing. It neatly divides the explanatory labor along traditional disciplinary lines and thus relieves members of each discipline of learning overly much about the other discipline. On one hand, psychologists are given the task of uncovering the functional organization of the mind without worrying about what neuroscientists do. On the other hand, neuroscientists are given the task of discovering neural mechanisms without having to think too hard about how the mind works. Everybody can do their job without getting in each other's way. Someday, if everything goes right, the functional analyses discovered by

---

Although we along with every other participant in this debate assume that functional systems are realized by mechanisms, some dualists disagree; they maintain that a functional system may be a non-physical, non-mechanistically-implemented system. We disregard this possibility on the usual grounds of causal closure of the physical and lack of an adequate account of the interaction between physical and non-physical properties. In any case, dualism is not our present target.

[10] Michael Strevens suggested to one of us in conversation that psychology and neuroscience may only constraint each other in this indirect way.

psychologists will turn out to be realized by the neural mechanisms discovered by neuroscientists. And yet, according to this autonomy thesis, neither the functional properties nor the structures place direct constraints on one another.

A number of philosophers have resisted autonomy understood as lack of direct constraints. Sometimes defenders of mechanistic explanation maintain that functional analysis—or "functional decomposition," as Bechtel and Richardson call it—is just a step towards mechanistic explanation (Bechtel and Richardson 1993, 89-90; Bechtel 2008, 136; Feest 2003). Furthermore, many argue that explanations at different mechanistic levels directly constrain one another (Bechtel and Mundale 1999, P.S. Churchland 1986, Craver 2007, Feest 2003, Keeley 2000, Shapiro 2004).

We agree with both of these points. But they do not go deep enough. The same authors who question autonomy seem to underestimate the role that distinctness plays in defenses of autonomy. Sometimes proponents of mechanistic explanation even vaguely hint or assume that functional analysis is *the same as* mechanistic explanation (e.g., Bechtel 2008, 140; Feest 2003; Glennan 2005), but they don't articulate and defend that thesis. So long as distinctness remains in place, defenders of autonomy have room to resist the mechanist's objections. Autonomists may insist that functional analysis, properly so called, is autonomous from mechanistic explanation after all. By arguing that functional analysis is actually a kind of mechanistic explanation, we get closer to the bottom of this dialectic. Functional analysis cannot be autonomous from mechanistic explanation because the former is just an elliptical form of the latter.

In the rest of this paper, we argue—along with others—that functional analysis and mechanistic explanation are not autonomous because they constrain each other; in addition, we argue that they can't possibly be autonomous in this sense because functional analysis is just a kind of mechanistic explanation. Functional properties are an undetachable aspect of mechanistic explanations. Any given explanatory text might accentuate the functional properties at the expense of the structural properties, but this is a difference of emphasis rather than difference in kind. The target of the description in each case is a mechanism.

In the next section, we introduce contemporary views of mechanistic explanation and show that mechanistic explanation is rich enough to incorporate the kinds of functional properties postulated by functional analysis. Then we argue that each kind of functional analysis is a mechanism sketch—an elliptical description of a mechanism.

3. Mechanistic Explanation

Mechanistic explanation is the explanation of the capacities (functions, behaviors, activities) of a system as a whole in terms of some of its components, their properties and capacities (including their functions, behaviors, or activities), and the way they are organized together (Bechtel and Richardson 1993, Machamer, Darden, and Craver 2000, Glennan 2002). Components have both functional properties—their activities or manifestations of their causal powers, dispositions, or capacities—and structural properties—including their location, shape, orientation, and the organization of their sub-components. Both functional and structural properties of components are aspects of mechanistic explanation.

Mechanistic explanation has also been called "system analysis," "componential analysis" (Cummins 1983, 28-9; 2000, 126), and "mechanistic analysis" (Fodor 1965). Constructing a mechanistic explanation requires decomposing the capacities of the whole mechanism into subcapacities (Bechtel and Richardson 1993). This is similar to task analysis, except that the subcapacities are assigned to structural components of a mechanism. The term "structural" does *not* imply that the components involved are neatly spatially localizable, have only one function, are stable and unchanging, or lack complex or dynamic feedback relations with other components. Indeed, a structural component might be so distributed and diffuse as to defy tidy structural description, though it no doubt has one if we had the time, knowledge, and patience to formulate it.

Mechanistic explanation relies on the identification of relevant components in the target mechanism (Craver 2007). Components are sometimes identified by their structural properties. For instance, some anatomical techniques are used to characterize different parts of the nervous system on the basis of the different kinds of neurons they contain and how such neurons are connected to one another. Brodmann decomposed the cortex into distinct structural regions by characterizing cyto-architectonic differences in different layers of the cortical parenchyma. Geneticists characterize the primary sequence of a gene. Such investigations are primarily directed at uncovering structural features rather than functional ones.

But anatomy alone cannot yield a mechanistic explanation—mechanistic explanation requires identifying the functional properties of the components. For example, case studies of brain damaged patients and functional magnetic resonance imaging are used to identify regions in the brain that contribute to the performance of some cognitive tasks. Such methods are crucial in part because they help to identify regions of the brain in which relevant structures for different cognitive functions might be found. They are also crucial to assigning functions to the different components of the mechanism. Each of these discoveries is a kind of progress in the search for neural mechanisms.

Functional properties are specified in terms of effects on some medium or component under certain conditions. Different structures and structure configurations have different functional properties. As a consequence, the presence of certain functional properties within a mechanism constrains the possible structures and configurations that might exhibit those properties. Likewise, the presence of certain structures and configurations within a mechanism constrains the possible functions that might be exhibited by those structures and configurations (cf. Sporns, Tononi, and Kötter 2005).

Mechanistic explanation is hierarchical in the sense that the functional properties (functions, capacities, activities) of components can also often be mechanistically explained. Each iteration in such a hierarchical decomposition adds another level of mechanisms to the hierarchy, with levels arranged in component/sub-component relationships and ultimately (if ever) bottoming out in components whose behavior has no mechanistic explanation.

Descriptions of mechanisms—mechanism schemas (Machamer, Darden, and Craver 2000) or models (Glennan 2005, Craver 2006)—can be more or less complete. Incomplete models—with gaps, question-marks, filler-terms, or hand-waving boxes and arrows—are mechanism sketches. Mechanism sketches are incomplete because they leave out crucial details about how the mechanism works. Sometimes a sketch provides just the right amount of explanatory information for a given context (classroom, courtroom, lab meeting, etc.). Furthermore, sketches are often useful guides to the future development of a mechanistic explanation. Yet there remains a sense in which mechanism sketches are incomplete or elliptical.

The common crux of mechanistic explanation, both in its current form and in forms stretching back through Descartes to Aristotle, is to reveal the causal structure of a system. Explanatory models are evaluated as good or bad to the extent that they capture, even dimly at times, aspects of that causal structure. Our argument is that the motivations guiding prominent accounts of functional analysis commit its defenders to precisely the same norms of explanation that mechanists embrace. One can embrace non-mechanistic forms of functional analysis only to the extent that one turns a blind eye to the very normative commitments that make functional analysis explanatory in the first place (for a different argument to this effect, see Kaplan and Craver forthcoming).

4. Task Analysis

A task analysis breaks a capacity of a system into a set of sub-capacities and specifies how the sub-capacities are (or may be) organized to yield the capacity to be explained. The organization may be simple temporal succession or more complex, as when the completion of subtask 2 after subtask 1 requires the system to go back to subtask 1.

Cummins calls the specification of the way sub-capacities are organized into capacities a "program" or "flow-chart" (Cummins 1975, 1983, 2000). The organization of capacities may also be specified by, say, a system of differential equations linking variables representing various sub-capacities. What matters is that one specifies how the various sub-capacities are combined or interact so that they give rise to the capacity of the system as whole.

Cummins remains explicitly ambiguous about whether the analyzing sub-capacities are assigned to the whole mechanism or to its components. The reason is that *at least in some cases*, the functional analysis of a capacity "seems to put no constraints at all on … componential analysis [i.e., mechanistic explanation]" (1983, p. 30). We discuss the different types of functional analysis separately. First, "functional analyses" that assign sub-capacities to *structural* components are mechanistic explanations (Craver 2001). Second, functional analyses that assign sub-capacities to *functional* components (black boxes) are boxological models (see Section 6). Finally, functional analyses that assign sub-capacities to the whole system rather than its components are task analyses. Cummins' examples of capacities subject to task analysis are assembly line production, multiplication, and cooking (Cummins 1975, 1983, 2000). Task analysis is the topic of this section.

Task analysis of a capacity is one step in its mechanistic explanation in the sense that it partitions the phenomenon to be explained into intelligible units or paragraphs of activity. For example, contemporary memory researchers partition semantic memory into encoding, storage, and retrieval processes, appeal to each of which will be required to explain performance on any memory task. Contra Cummins, the partition of the phenomenon places direct constraints on components, their functions, and their organization. For if a sub-capacity is a genuinely explanatory part of the whole capacity, as opposed to an arbitrary partition (a mere piece or temporal slice), it must be exhibited by specific components or specific configurations of components. In the systems with which psychologists and neuroscientists are concerned, the sub-capacities are not ontologically primitive; they belong to structures and their configurations. The systems have the capacities they have in virtue of their components and organization.

Consider Cummins's examples: stirring (a sub-capacity needed in cooking) is the manifestation of (parts of) the cook's locomotive system coupled with an appropriate stirring tool as they are driven by a specific motor program; multiplying single-digit numbers (a sub-capacity needed in multiplying multiple digit numbers) is the manifestation of a memorized look-up table in cooperation with other parts of the cognitive system; and assembly line production requires different machines and operators with different skills at different stages of production. Likewise, the correctness of the above-mentioned task analysis of memory into encoding, storage, and retrieval depends on whether there are components that encode, store, and retrieve memories.

Task analysis is constrained, in turn, by the available components and modes of organization. If the study of brain mechanisms forces us to lump, split, eliminate or otherwise rethink any of these sub-capacities, the functional analysis of memory will have to change (cf. Craver 2004). If the cook lacks a stirring tool but still manages to combine ingredients thoroughly, we should expect that the mixture has been achieved by other means, such as shaking. If someone doesn't remember the product of two single digit numbers, she may have to do successive sums to figure it out. The moral is that if the components predicted by a given task analysis are not there or are not functioning properly, you must rule out that task analysis in favor of another for the case in question (cf. Craver and Darden 2001).

In summary, a task analysis is a mechanism sketch in which the capacity to be explained is articulated into sub-capacities, and most of the information about components is omitted. Nevertheless, the sub-capacities do place direct constraints on which components can engage in those capacities. For each sub-capacity, we expect a structure or configuration of structures that has that capacity. This guiding image underlies the very idea that these are explanations, that they reveal the causal structure of the system. If the connection between analyzing tasks and components is severed completely, then there is no clear sense in which the analyzing sub-capacities are aspects of the *actual* causal structure of the system as opposed to arbitrary partitions of the system's capacities or merely possible causal structures. Indeed, components often call attention to themselves as components only once it is possible to see them as performing what can be taken as a unified function.

So task analysis specifies sub-capacities to be explained mechanistically, places direct constraints on the kinds of components and modes of organization that might be employed in the mechanistic explanation, and is in turn constrained by the components, functional properties, and kinds of organization that are available.[11]

At this point, a defender of the distinctness and autonomy of functional analysis may object that although many task analyses are first-stage mechanistic theories as we maintain, the really interesting cases, including the ones most pertinent to cognitive phenomena, are not. The most influential putative example of such is the *general*

---

[11] This account of task analysis is borne out by the historical evolution of task analysis techniques in psychology. Psychologists developed several techniques for analyzing complex behavioral tasks and determine how they can be accomplished efficiently. Over time, these techniques evolved from purely behavioral techniques, in which the sole purpose is to analyze a task or behavior into a series of operations, into cognitive task analysis, which also aims at capturing the agents' cognitive states and their role in guiding their behaviors. To capture the role of cognitive states, agents ought to be decomposed into their components (Crandall, Klein and Hoffman, 2006, 98). The motivation for such a shift is precisely to capture more accurately the way agents solve problems. Focusing solely on sequences of operations has proved less effective than analyzing also the agents' underlying cognitive systems, including their components. As the perspective we are defending would predict, task analysis in psychology has evolved from a technique of behavioral analysis, closer to task analysis as conceived by Cummins, towards a more mechanistic enterprise.

purpose computer. General purpose computers can do an indefinite number of things depending on how they are programmed. Thus, one might think, a task analysis of a general purpose computer's capacities places no direct constraints on its structural components and the structural components place no direct constraints on task analysis, because the components are the same regardless of which capacity a computer exhibits. The only constraint is indirect: the computer must be general purpose, so it must have general purpose components. By extension, if the same kind of autonomous task analysis applies to human behavior, then the type of task analysis with which the received view is concerned is not a mechanism sketch.

This objection makes an important point but draws the wrong conclusion. True, general purpose computers are different from most other systems precisely because they can do so many things depending on how they are programmed. But general purpose computers are still mechanisms, and the explanation of their behavior is still mechanistic (Piccinini 2007, 2008). Furthermore, the task analysis of a general purpose computer does place direct constraints on its mechanistic explanation and vice versa; in fact, even the task analysis of a general purpose computer is just an elliptical mechanistic explanation.

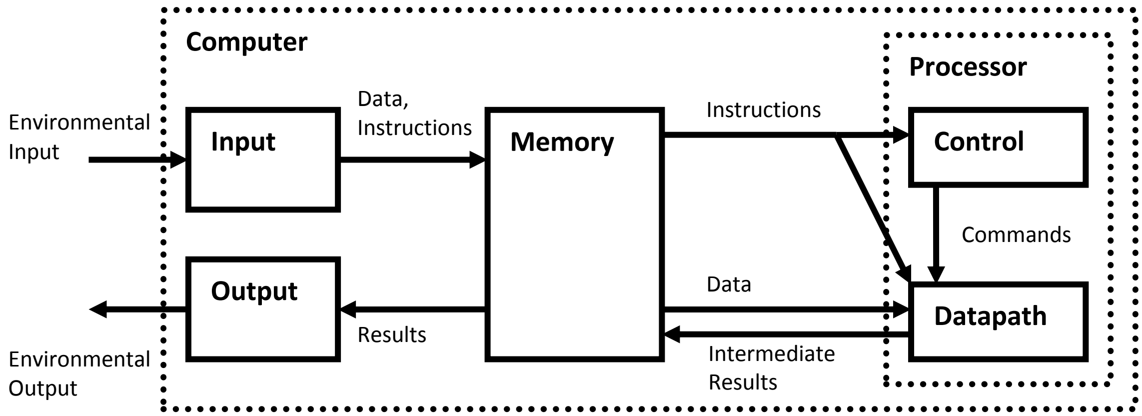

Figure 1: Functional organization of a general purpose digital computer.

To begin with, the explanation of the capacity to execute programs is mechanistic: a processor executes instructions over data that are fed to it and returns results to other components (Figure 1). In addition, the explanation of any given computer behavior is also mechanistic: the processor executes *these particular instructions*, and it is *their* execution that results in the behavior. Because the program execution feature is always the same, in many contexts it is appropriate to omit that part of the explanation. But the result is not a non-mechanistic task analysis; it is, again, an elliptical one. It is a task analysis in which most of the mechanism is left implicit—the only part that is made explicit is the executed program.

Notice that for this type of explanation to be appropriate for human beings, they must turn out to contain general purpose computers; thus, they must turn out to contain processors capable of executing the right kind of instructions plus memory components to store data, instructions, and results. Whether human brains contain this kind of mechanism is an empirical question, and it can only be resolved by investigating whether brains have this kind of organization. If the analogy between brains and general purpose computers is more than just a metaphor, it has to make substantive claims about brain mechanisms. Those claims might turn out to be false, in which case we need to revise our understanding of brains and how they work.

A defender of the received view may reply as follows. Granted, whether a system is a general purpose computer is a matter of which mechanisms it contains. But *if* we can assume that a system is a general purpose computer, *then* we can give task analyses of its behavior that say nothing about its mechanisms. For when we describe the program executed by the computer, we are not describing any components or aspects of the components.[12] We are abstracting away from the mechanisms.

This objection misconstrues the role programs play in the explanation of computer behavior. There are two cases. First, if a computer is hardwired to perform a certain computation, the computer's behavior may still be described by a flow-chart or

---

[12] Cf. Cummins on programs: "programs aren't causes but abstract objects or play-by-play accounts" (Cummins 1983, 34).

program.[13] In the case of hardwired computations, it is true that a program is not a component of the computer. But it is false that the program description is independent of the description of the machine's components. In fact, the program describes precisely the activity of the special purpose circuit that is hardwired to generate the relevant computation. A circuit is a structural component. Therefore, a task analysis of a computer that is hardwired to perform a computation is a precise description of a specific structural component. Incidentally, an analogous point applies to connectionist systems that come to implement a task analysis through the adjustment of weights between nodes.

The second case involves those computers that produce a behavior by *executing* a program—that is, by being caused to produce the behavior by the presence of the program within them. Here, the program itself is a stable state of one or more computer components. Programs are typically stored in memory components and sent to processors one instruction at a time. If they weren't physically present within computers, digital computers could not execute them and would not generate the relevant behaviors. Thus programs are a necessary feature of the mechanistic explanation of computers' behaviors; any task analysis of a computer in terms of its program is an elliptical mechanism sketch that provides only the program and elides the

[13] In these cases, we find it misleading to say, as Cummins and others do, that the computer is *executing* the program, because the program is not an entity in its own right that plays a causal role. The same point applies to computers that are programmed by rearranging the connections between their components.

rest of the mechanism (including, typically, the subsidiary programs that translate high level code into machine executable code).

A final objection might be that some computational models focus on the flow of information through a system rather than the mechanisms that process the information (cf. Shagrir 2006, 2010). In such cases, nothing is added to the explanation by fleshing out the details of how the information is represented and processed. Certainly, many computational explanations in psychology and neuroscience have this form. Our point, however, is that such descriptions of a system place direct constraints on any structures that can possibly process such information—on how the different states of the system can be constructed, combined, and manipulated—and are in turn constrained by the structures to be found in the system. It is, after all, an empirical matter whether the brain has structural components that satisfy a given informational description, that is, whether the neuronal structures in question can sustain the information processing that the model posits (under ecologically and physiologically relevant conditions). If they cannot, then the model is a how-possibly model and does not describe how the system actually works. Computational or informational explanations are still tethered to structural facts about the implementing system. Once we know the information processing task, we might think that details about how the information is encoded and manipulated are no longer relevant, and in some explanatory contexts this is true, but details about how the information is encoded and manipulated are, in fact, essential to confirm our hypotheses about information processing.

Our reason for thinking that task analysis boils down to (elliptical) mechanistic explanation might be summarized as follows. Either our opponent accepts that there is a unique correct task analysis for a given system capacity at a given time or she does not. If explanations must be true, the idea of a unique correct task analysis for a given system capacity amounts to the idea that there are units of activity between inputs and outputs that transform the relevant inputs (and internal states) into outputs. In real psychobiological systems, flesh and blood lie between the input and the output, and it has been organized through evolution and development to exhibit that behavior. Whether the flesh and blood satisfies the given task analysis depends on whether the flesh and blood contains units of structure that complete each of the tasks. If no such units of structure can be identified in the system, then the task analysis cannot be correct.[14]

Our imagined opponent might thus prefer to deny that there is a unique correct task analysis for a given system behavior. Cummins sometimes seems to embrace this view (see Cummins 1983, 43). Perhaps our opponent maintains that the beauty of functional analysis is that it allows us to rise above the gory details, details that often vary from species to species, individual to individual, and time to time, and thereby to "capture" features of the causal structure of the system that are invisible when we focus on the microstructural details. She might think that any model that describes correctly and

[14] Lest there be some confusion on this point, we emphasize again that components need not be neatly localizable, visible, or spatially contained within a well defined area. Any robustly detectable configuration of structural properties might count as a component.

compactly the behavior of the system is equally explanatory of the system's behavior. Or she might think that higher-level descriptions can render the phenomenon intelligible even when they do not have any tidy echo in microstructural details.

We agree that models can be predictively adequate and intellectually satisfying even when they fail to describe mechanisms. Our opponent, however, maintains that predictive adequacy and/or intelligibility are enough for explanation. But this view of explanation is hard to defend. In fact, most philosophers of science agree that neither predictive adequacy nor intelligibility is sufficient for explanation. One can predict an empty gas tank without knowing how it came to be empty. And anyone who has ever misunderstood how something works is familiar with the idea that intelligible explanations can be terrible explanations. Some explanations make phenomena intelligible, but not all intelligible models are explanatory. In short, to give up on the idea that there is a uniquely correct explanation, and to allow that any predictively adequate and/or intelligible model is explanatory, is essentially to give up on the idea that there is something distinctive about explanatory knowledge. No advocate of functional analysis should want to give that up.

Either there is a uniquely correct task analysis, in which case the underlying mechanism must have components corresponding to the sub-tasks in the analysis, or there is not a uniquely correct task analysis, in which case task analysis has been severed from the actual causal structure of the system and does not count as explanation. In short, either task analysis is an elliptical form of mechanistic explanation or it is no explanation at all.

5. Functional Analysis by Internal States

The capacities of a system, especially cognitive capacities, are sometimes said to be explained by the system's internal states (and internal processes, defined as changes of internal states). Common examples of internal states include propositional attitudes, such as beliefs and desires, and sensations, such as pain. In an important strand of the philosophy of psychology, internal states are held to be functional states, namely, states that are individuated by their relations to inputs, outputs, and other internal states (Putnam 1960, 1967a, b; Fodor 1965, 1968). Within this tradition, the functional relations between inputs, outputs, and internal states are said to constitute the functional organization of the system, on which the functional analysis of the capacities of the system is based.[15]

This account of mental states as internal functional states has been challenged. But our concern here is not whether the account is adequate. Our concern is, rather, how functional analysis by internal functional states relates to mechanistic explanation.[16] In

[15] Although this functionalist account is often said to be neutral between physicalism and dualism, this is an oversimplification. The internal states interact with the inputs and outputs of the system, which are physical. Thus, if the internal states were non-physical, they must still be able to interact with the physical inputs and outputs. This violates the causal closure of the physical and requires an account of the interaction between physical and the non-physical states. So, while a version of functionalism may be *defined* to be compatible with dualism, any metaphysically respectable version of functionalism should be physicalistic.

[16] Notice that insofar as we are dealing with explanations of cognitive capacities, we are focusing on states that are internal to the cognitive system, whether or not the system's boundaries coincide with the body or nervous system of the organism.

order to assess functional analysis by internal states, it is necessary to ask the further question: In what sense are such functional states *internal*?

The notion of state may be taken as primitive or analyzed as the possession of a property at a time. Either way, in general, there is no obvious sense in which a state of a system, per se, is internal to the system. For example, consider the distinction between the solid, liquid, and gaseous states of substances such as water. There is no interesting sense in which the state of being liquid, solid, or gaseous is internal to samples of water or any other substance. Of course, the scientific explanation for why, at certain temperatures, water is solid, liquid, or gaseous involves the components of water (H_2O molecules) and their temperature-related state, which, together with other properties of the molecules (such as their shape), generates the molecular organization that constitutes the relevant global state of water samples. In this explanation we find a useful notion of internal state because individual water molecules are contained within the admittedly imprecise spatial boundaries of populations of water molecules.

The moral is this: in general, a system's states (simpliciter) are not internal in any interesting sense; they are global, system-level states. But there is something interestingly internal to the state of the *components* of a system, which play a role in explaining the global state of a system (as well as its behavior). This is because the components are *inside* the system. Are the internal states invoked in functional analysis system-level states or states of components? The qualifier "internal" suggests the latter.[17] Here is why the internal states postulated by a functional analysis must be states of the system's components.

Functional analysis by internal states postulates a system of multiple states. Such states are capable of interacting with inputs, outputs, and each other, in order to produce novel states and outputs from previous states and inputs. Notice three features of these systems. First, it must be possible for many states to occur at the same time. Second, inputs and outputs enter and exit through specific components of the system. Third, inputs and outputs are complex configurations of different physical media, such as light waves, sound waves, chemical substances, and bodily motions.

The only known way to construct a system of states that can occur at the same time and mediate between such inputs and outputs is to transduce the different kinds of input into a common medium, different configurations of which are the different states (e.g., configurations of letters from the same alphabet or patterns of activation of mutually connected neurons). For different configurations of the same medium to be present at the same time, they must be possessed by different components. The agents in a causal

[17] In any case, etymology does support our conclusion. Putnam imported the notion of an internal state into the philosophy of psychology as part of his (1960) analogy between mental states and Turing machine states. Turing described digital computers as having "internal states" (Turing 1950) that belong to the read-write head, which is the active component of Turing machines. In the case of digital computers, internal states are states of an internal component of the computer, such as the memory or processor. Initially, Putnam did not call states "internal" but "logical" (1960, 1967a), and then he called them "functional" states (1967b). But Fodor called them "internal" states (1965; cf. also 1968a, Block and Fodor 1972, and Piccinini 2004). Via Putnam and Fodor, Turing's phrase "internal state" has become the established way to talk about functional states. As we have seen, it originally referred to the states of a component (of a computing machine).

interaction have to be distinct from one another. Furthermore, such components must be able to create the relevant configurations of the medium, distinguish between them, and interact with each other so as to produce the relevant subsequent states and outputs from the previous states and inputs. Thus, functional analysis by internal states requires that the states belong to some of the system's components and constrains the properties of the components.

A system can surely possess different global states at the same time, e.g., a color, a speed, and a temperature. Such global states can affect each other as well as the behavior of the system. For instance, a system's color influences heat absorption, which affects temperature, which in turn affects heat dissipation. But global states can only influence global variables—they cannot mediate between complex configurations of different physical media coming through different sensory systems and generate the specialized configurations of outputs coming from a motor system. Heat absorption and dissipation are not complex configurations of a physical medium—they are global variables themselves. Global variables such as color and temperature can affect other global variables such as heat absorption and dissipation—they cannot transform, say, a specific pattern of retinal stimulation into a specific pattern of muscle contractions. Or at any rate, no one has ever begun to show that they can.

One might insist, however, that functional analysis in terms of functional states makes no reference (directly or indirectly) to components, and so need not be a mechanism sketch. The goal of an explanation is to capture in a model how a system behaves;

models need not describe components in order to capture how a system behaves; models need not describe components in order to explain.

The problem with this view, as with the analogous conception of task analysis above, is that it confuses explaining with modeling. A resounding lesson of 50 years of sustained discussion of the nature of scientific explanation is that not all phenomenally and predictively adequate models are explanations. One can construct models that predict phenomena on the basis of their correlations (as barometers predict but do not explain storms), regular temporal successions (national anthems precede but do not explain kickoffs), and effects (as fevers predict but do not explain infections). Furthermore, there is a fundamental distinction between redescribing a phenomenon (even in law-like statements) and explaining the phenomenon. Snell's law predicts how light will bend as it passes from one medium to another, but it does not explain why light bends as it does. One might explain that the light bent because it passed from one medium to another, of course. But that is an etiological explanation of some light-bending events, not a constitutive explanation of why light bends when it passes between different media.

Finally, one can build predictively adequate models that contain arbitrarily large amounts of superfluous (i.e., nonexplanatory) detail. Explanations are framed by considerations of explanatory relevance. If functional analysis by internal states is watered down to the point that it no longer makes any commitments to the behavior of components, then it is no longer possible to distinguish explanations from merely predictively adequate models and phenomenal descriptions of the system's behavior. Nor does such a modeling strategy tell us how to eliminate irrelevant information from our explanations—a crucial explanatory endeavor. In short, "explanation" of this sort is not worthy of the name.

In conclusion, "internal" states either are not really internal, in which case they constitute a system-level explanandum for a mechanistic explanation, or they are internal in the sense of being states of components. As we have seen, there are two ways to think of components. On one hand are structural components. In this case, functional analysis by internal states is a promissory note on (a sketch of) a mechanistic explanation. The analysis postulates states of some structural components, to be identified by a complete mechanistic explanation.[18] On the other hand, components may also be functionally individuated components or black boxes. (For instance, the read-write head of Turing machines is a paradigmatic example of a black box.) When components are construed as black boxes, functional analysis by internal states becomes boxology, to which we now turn our attention.

---

[18] As we have seen, Fodor says that functional analyses give no information about the mechanism underlying these states (1965, 177), but at least, they entail that there are components capable of bearing those states and capable of affecting each other so as to generate the relevant changes of states:

> [I]t is sufficient to disconfirm a functional account of the behaviour of an organism to show that its nervous system is incapable of assuming states manifesting the functional characteristics that account requires… it is clearly good strategy for the psychologist to construct such theories in awareness of the best estimates of what the neurological facts are likely to be (Fodor 1965, 176).

Much of what Fodor says in his early works on functional analysis is in line with our present argument and goes against the autonomy assumption that Fodor later defended. For simplicity, in the main text we are assimilating Fodor's early (anti-autonomy) writings to his later (pro-autonomy) writings. In any event, even Fodor's early writings fall short of pointing out that functional analyses are mechanism sketches.

6. Boxology

Black boxes are components individuated by the outputs they produce under certain input conditions. In this sense, they are functionally individuated components. In another important strand of the philosophy of psychology, the capacities of a system are said to be functionally explained by appropriately connected black boxes. For example, Fodor distinguishes the functional identification of components from their structural identification (Fodor 1965, 1968a; for a similar distinction, see also Harman 1988, 235).[19] Black boxes are explicitly *internal*—spatially contained within the system. Proponents of boxology appear to believe that capacities can be satisfactorily explained in terms of black boxes, without identifying the structural components that implement the black boxes. What proponents of this type of functional analysis fail to notice is that functional and structural properties of components are interdependent: both are necessary, mutually constraining aspects of a mechanistic explanation. On one hand, the functional properties of a black box constrain the range of structural components that can exhibit those functional properties. On the other hand, a set of structural components can only exhibit certain functional properties and not others.

Consider Fodor's example of the camshaft. Internal combustion engines, Fodor reminds us, contain valves that let fuel into the pistons. For fuel to be let in, the valves need to be lifted, and for valves to be lifted, there must be something that lifts the valves. So

[19] E.g.: "In functional analysis, one asks about a part of a mechanism what role it plays in the activities characteristic of the mechanism as a whole" (Fodor 1965, p. 177).

here is a job description—valve lifting—that can be used to specify what a component of an engine must do for the engine to function. It is a functional description, not a structural one, because it says nothing about the structural properties of the components that fulfill that function, or how they manage to fulfill it. There may be indefinitely many ways to lift valves; as long as something does, it qualifies as a valve lifter. (Hence, multiple realizability.)

What are the components that normally function as valve lifters in internal combustion engines? Camshafts. This is now a structural description, referring to a kind of component individuated by its shape and other structural properties. So there are two independent kinds of description, Fodor concludes, in terms of which the capacities of a system can be explained. Some descriptions are functional and some are structural. Since Fodor maintains that these descriptions are independent, there is a kind of explanation—boxology—that is autonomous from structural descriptions. Functional descriptions belong in boxological models, whereas structural descriptions belong in mechanistic explanations. Or so Fodor maintains. But is it really so?

In the actual "functional analysis" of a mechanism, such as an engine, the functions are specified in terms of physical effects either on a physical medium or on other components, both of which are structurally individuated. Valve lifting means physically lifting a valve. The functional description "valve lifter" contains two terms. The first, "valve," refers to a kind of component, whereas the second, "lifter," refers to a capacity. Neither of these appears to be independent of structural considerations. Lifting is a physical activity: for *x* to lift *y*, *x* must exert an appropriate physical force on *y* in the relevant direction. The notion of valve is both functional and structural. In this context, the relevant sense is at least partially structural, for nothing could be a valve in the sense relevant to valve lifting unless it had weight that needs to be lifted for it to act as a valve. As a consequence, the "valve lifter" job description puts three mechanistic constraints on explanation: first, there must be valves (a type of structural component) to be lifted; second, lifting (a type of structurally individuated capacity) must be exerted on the valves; and third, there must be valve lifters (another type of component) to do the lifting. For something to be a valve lifter in the relevant respect, it must be able to exert an appropriate physical force on a component with certain structural characteristics in the relevant direction. This is not to say that only camshafts can act as valve lifters. Multiple realizability stands. But it is to say that all valve lifters suitable to be used in an internal combustion engine share certain structural properties with camshafts.

This point generalizes. There is no such thing as a purely functional analysis of the capacity of an engine to generate motive power. Any attempt to specify the nature of a component purely functionally, in terms of what it is for, runs into the fact that the component's function is to interact with other components to exhibit certain physical capacities, and the specification of the other components and activities is inevitably infected by structural considerations. Here is why. The inputs (flammable fuel, igniting sparks) and the outputs (motive power, exhaust) of an internal combustion engine are concrete physical media that are structurally individuated. Anything that turns structurally individuated inputs into structurally individuated outputs must possess appropriate physical causal powers—powers that turn those inputs into those outputs. (This, by the way, does not erase multiple realizability: there are still many ways to build an internal combustion engine.)

What about boxological models in psychology and neuroscience? A boxologist committed to autonomy may suggest that while our assimilation of boxological models to mechanism sketches is viable in most domains, including internal combustion engines, it does not apply to computing systems. In this special domain, our boxologist continues, the inputs and outputs can be specified independently of their physical implementation and black boxes need not even correspond to concrete components.

In particular, Marr (1982) is often interpreted as arguing that there are three autonomous "levels" of explanation in cognitive science: a "computational level," an "algorithmic level," and an "implementational level". According to Marr, the computational level describes the computational task, the algorithmic level describes the representations and representational manipulations by which the task is solved, and the implementational level describes the mechanism that carries out the algorithm. Marr's computational and algorithmic levels may be seen as describing black boxes independently of their implementation.

Since the functional properties of black boxes are specified in terms of their inputs and outputs (plus the algorithm, perhaps), the black boxes can be specified independently of their physical implementation. Thus, at least in the case of computing systems, which presumably include cognitive systems, boxology does not reduce to mechanistic explanation.

This reply draws an incorrect conclusion from two correct observations. The correct observations are that black boxes may not correspond one-to-one to structural components and that the inputs and outputs (and algorithms) of a computing system can be specified independently of the physical medium in which the inputs and outputs are implemented. As a result, the same computations defined over the same computational vehicles can be implemented in various—mechanical, electro-mechanical, electronic, etc.—physical media. Indeed, some physical properties of the media are irrelevant to whether they implement a certain computation or another.

But it doesn't follow that computational inputs and outputs put no direct constraints on their physical implementation. In fact, any physical medium that implements a certain computation must possess appropriate physical degrees of freedom that result in the differentiation between the relevant computational vehicles. Furthermore, any component that processes computational vehicles must be able to reliably discriminate between tokens of the relevant types so as to process them correctly. Finally, any components that implement a particular algorithm must exhibit the relevant kinds of operations in the appropriate sequence. The operations in question are different depending on how black boxes map onto structural components.

Consider a staple of functionalist philosophy of psychology: belief and desire boxes. Anyone who approves of such talk knows that belief and desire boxes need not be two separate memory components. True, but this doesn't entail lack of direct constraints between functional properties and structural components. For the alternative means of implementing belief and desire boxes is to store beliefs and desires in one and the same memory component, while setting up the memory and processor(s) so that they can keep track of which representations are beliefs and which are desires. This may be done by adding an attitude-relative index to the representations or by keeping lists of memory registers. However it is done, the distinction between beliefs and desires constrains the mechanism: the mechanism must distinguish between the two types of representation and process them accordingly (if the organism is to exhibit relevant behavior). And the mechanism constrains the functional analysis: the representational format and algorithm will vary depending on how the distinction between beliefs and desires is handled by the mechanism, including whether each type of representation has a dedicated memory component. Thus, even though the decomposition into black boxes may not correspond one-to-one to the decomposition into concrete components, it still constrains the properties of concrete components.

The moral is that computing systems are indeed different from most other functionally organized systems, in that their computational behavior can be mechanistically explained without specifying the physical medium of implementation other than by specifying which degrees of freedom it must possess. But such an explanation is still mechanistic: it specifies the type of vehicle being processed (digital, analog, or what have you) as well as the structural components that do the processing, their organization, and the functions they compute. So computational explanations are mechanistic too (Piccinini 2007, Piccinini and Scarantino 2010).

What about Marr's computational and algorithmic "levels"? We should not be misled by Marr's terminological choices. His "levels" are not levels of mechanism because they do not describe component/sub-component relations. (The algorithm is not a component of the computation, and the implementation is not a component of the algorithm.) The computational and algorithmic levels are mechanism sketches. The "computational level" is a description of the mechanism's task, possibly including a task analysis, whereas the "algorithmic level" is a description of the computational vehicles and processes that manipulate the vehicles. All of the above—task, vehicles, and computational processes—constrain the range of components that can be in play and are constrained in turn by the available components. Contrary to the autonomist interpretation of Marr, his "levels" are just different aspects of the same mechanistic explanation.

So black boxes are placeholders for structural components (with arrows indicating input-output relations between components) or sub-capacities (with arrows indicating causal relations between processes) in a mechanism. Boxology is not distinct from mechanistic explanation. Rather it is a first step toward the decomposition of a system into its structural components. Computational and information processing explanations often work by abstracting away from many of the implementing details. But that's how mechanistic explanation generally works; it focuses on the mechanistic level most relevant to explaining a behavior while abstracting away from the mechanistic levels below. Whether a system implements a given computation still depends on its structural features.

7. An Example of Functional Analysis

"Neurotransmitter" is a functional term, and it is a well defined one. When someone asks whether a molecule acts as a neurotransmitter at a synapse there is no ambiguity as to what question is being asked or what kind of evidence would be required to answer it. Table 1 lists six criteria—repeated in most if not all standard neuroscience texts (see, e.g. Kandel et al. 2000, 280-281; Shepherd 1994, 160; Cooper, et. al. 1996, 4)—for establishing that a molecule is a neurotransmitter.

Table 1. Six traditional criteria for identifying a neurotransmitter.

1. The chemical must be present in the presynaptic terminal.

2. The chemical must be released by the presynaptic terminal in amounts sufficient to exert its supposed action on the post-synaptic neuron (or organ). Release should be dependent upon inward calcium current and the degree of depolarization of the axon terminal during the action potential.

3. Exogenous application of the chemical substance in concentrations reasonably close to those found endogenously must mimic exactly the effect of endogenously released neurotransmitter.

4. The chemical must be synthesized in the presynaptic cell.

5. There must be some means of removing the chemical from the site of action (the synaptic cleft).

6. The effects of the putative neurotransmitter should be mimicked by known pharmacological agonists and should be blocked by known antagonists for that neurotransmitter.

Although each of these criteria is violated for some known neurotransmitters (especially amino acid transmitters like glutamate), they are nonetheless prototypical. These criteria are clearly designed to show that the putative neurotransmitter is organized within the mechanisms of chemical neurotransmission shown in Figure 2. We will see that attributing a function amounts to showing how some item fits into a mechanism. In order to understand how criteria 1-6 achieve that objective, we need first to review some of the basic mechanisms of chemical neurotransmission.

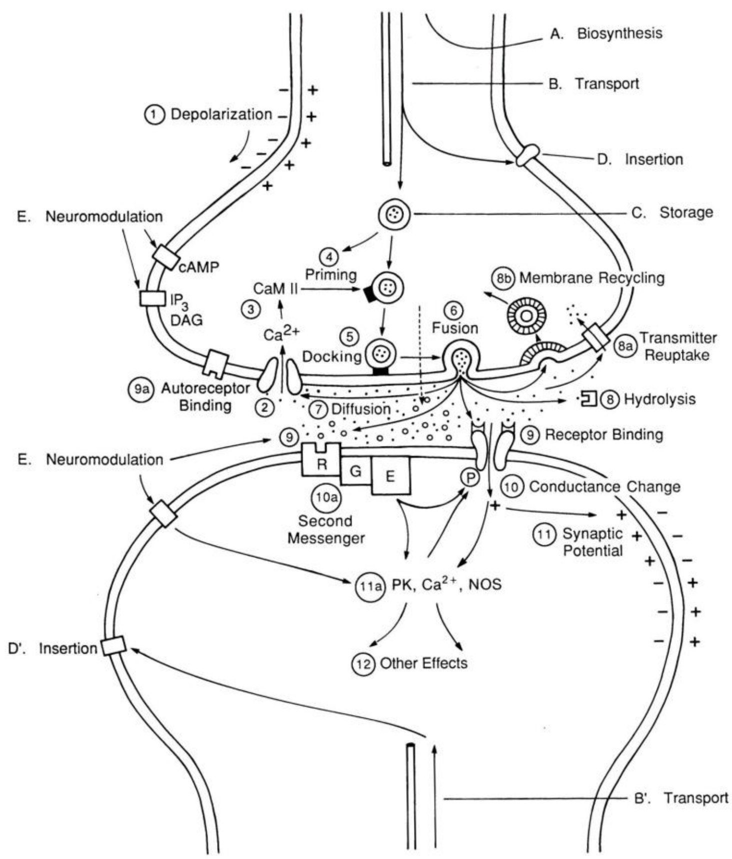

Figure 2. The mechanisms of chemical neurotransmission.

Neurons communicate with one another by passing neurotransmitters across synapses, gaps between a pre-synaptic axon (shown at the top of Figure 2) and a post-synaptic neuron (shown at the bottom of Figure 2). The model for a standard form of chemical neurotransmission begins with the depolarization of the axon terminal following the arrival of an action potential (1). When the axon depolarizes, voltage sensitive calcium (Ca^{2+}) channels open (2), allowing Ca^{2+} to diffuse into the cell. This sudden rise in intracellular calcium concentrations activate Ca^{2+}/calmodulin protein kinase, which can then prime the vesicles containing neurotransmitter (4). This neurotransmitter may have been synthesized (A) in the cell body and transported (B) to the axon terminal, or they may be stored in neurotransmitters in the terminal itself (C). Once primed, vesicles can then dock to the axon wall (5) in such a way that the vesicles can fuse (6) to the membrane and spill their transmitters into the synaptic cleft. These transmitters diffuse (7) across the synapse, where they bind to post-synaptic receptors, initiating either chemical or electrical effects in the post-synaptic cell. This cartoon of chemical neurotransmission is a mechanistic model, involving entities (ions, neurotransmitters, vesicles, membranes) and activities (depolarizing, diffusing, priming, docking, fusing) organized together so that they do something—in this case, reliably preserve a signal across the space between cells.

Identifying something as a neurotransmitter involves showing that it is organized within this set of mechanisms. Each of the criteria in Table 1 is designed to show that the putative neurotransmitter is organized—spatially, temporally, actively and quantitatively—into the mechanisms of chemical transmission. Spatially, the transmitter has to be located in the presynaptic neuron and identified as contained within vesicles. Temporally, if the synapse has to convey signals passed by electrical signals that are hundreds of milliseconds apart, there needs to be some means for removing the transmitter from the cleft after each episode of release. The transmitter has to be actively integrated within the mechanisms of synthesis of the presynaptic cell (again this is sometimes violated) and with the receptor mechanisms of the post-synaptic cell (namely, by showing that the post-synaptic cell responds to the presence of the molecule or by showing the putative post-synaptic effect of the chemical substance can be mimicked by applying pharmacological relatives of the transmitter and should be blocked by pharmacological antagonists for the receptors). And quantitatively the molecule's release must be correlated with activation of the pre-synaptic cell, and it must be released in sufficient quantity to affect the post-synaptic cell.

Some of these criteria (such as localization, vesicular packaging, and concentration) make direct appeal to structural features of the synapse. Others involve existential claims about inactivating molecules at the same location and causal claims about inhibitors and agonists. None of these features is independent of the causal details concerning the rest of the mechanism. The example thus illustrates how the functional language at lower-levels of neuroscience involves implicit commitments to structural facts about the mechanism in question. The same, we claim, extends all the way to the highest levels of the neuroscience hierarchy.

Functional analysis borrows its explanatory legitimacy from the idea that functional explanations (functional explanatory texts) capture something of the causal structure of a system. It has been understood as a doctrine according to which mental states (and other "high-level properties") are ultimately to be analyzed in terms of the commerce between them (as well as inputs and outputs). Haugeland (1998) likens the situation to recent changes in the structure of departments within a company. As communication technologies continue to expand, physical locations have become much less significant than are facts about how the lines of communication flow: of who reports to whom, of what tasks are divided among the nodes in this network, of the sorts of communication passed along the different interfaces. There are, to be sure, facts about location, structure, etc. But those facts turn out not to be so important if you want to understand how the department works. Is Larry's office next to Bob's, or is it in Saigon? It just doesn't matter so long as their communication can happen in the time required. Do they communicate by speaking, by telephone, by email, by instant messaging? Again, that just doesn't matter. Or so we might think.

Given that psychological systems are in fact implemented in biological systems, and that such systems are more or less precisely replicated through reproduction, evolution, and development, there are frozen structural constraints on the mechanisms that do, as a matter of fact, implement behavioral and cognitive functions. Learning about components allows one to get the right functional decomposition by ruling out functional decompositions that are incompatible with the known structural details (just as Larry and Bob could not run a tech firm between St. Louis and Saigon by Pony

Express). Explanations that can accommodate these details about underlying mechanisms have faced severe tests not faced by models that merely accommodate the phenomenon, and learning about such details is clearly of central importance in the discovery process. In the case of the neurotransmitter, all of the criteria seem to address exactly this point. So the search for mechanistic details is crucial to the process of sorting correct from incorrect functional explanations. If functional analysis is treated as different in kind from mechanistic explanation, it is hard to see how research into mechanistic details can play this role.

A second point is that explanations that capture these mechanistic details are deeper than those that do not. To accept as an explanation something that need not correspond with further levels of detail about how the system is in fact implemented is to accept that the explanations simply end at that point. Psychology ought to construct its explanations in a way that affords deeper filling in. That allows for progress in explanatory depth. Progress in explanatory depth has two virtues beyond the joy of understanding (which, as noted above, sometimes accompanies knowledge of a mechanism). First it allows one to expand the range of phenomena that the model can save. Compare two models, one that characterizes things without mentioning the structural details, and one that includes, in addition, structural facts about component parts. The latter allows us to make accurate predictions about how the mechanism will behave under a wider variety of variations in background conditions: the effective functioning of a company in a blackout does depend on whether Larry and Bob communicate in person or by instant messaging. Second, and related, knowledge of the underlying components and the structural constraints on their activities affords more opportunities for the restoration of function and the prevention of disease.

Our point is thus conceptual and pragmatic. On the conceptual side, we emphasize that functional analysis and mechanistic explanation are inextricably linked: structural descriptions constrain the space of plausible functional descriptions, and functional descriptions are elliptical mechanistic descriptions. The idea that functional description is somehow autonomous from details about mechanisms involves a fundamental misunderstanding of the nature of functional attribution in sciences like cognitive neuroscience. On the pragmatic side, the demand that explanations satisfy mechanistic constraints leads us to produce better, deeper, better confirmed, and more useful descriptions of the system at hand than we would produce if we allowed ourselves to be satisfied with any empirically adequate boxological models. For these reasons, full-blown mechanistic models are to be preferred.

In conclusion, there is no functional analysis that is distinct and autonomous from mechanistic explanation because to describe an item functionally is, ipso facto, to describe its contribution to a mechanism. Furthermore, a full blown mechanistic explanation describes both the functional and the structural properties of the mechanism; any constitutive explanation that omits the structural properties in favor of the functional ones is not a non-mechanistic explanation but an elliptical mechanistic explanation, at least if it is in fact worthy of the title "explanation" in this domain.[20]

8. Conclusion

Functional analysis of a system's capacities provides a sketch of a mechanistic explanation. If the functional analysis is just the explanation of the capacities of the system in terms of the system's sub-capacities, this is an articulation of the phenomenon to be mechanistically explained that points in the direction of components possessing the sub-capacities. If the functional analysis appeals to internal states, these are states of internal components, which need to be identified by a complete mechanistic explanation. Finally, a functional analysis may appeal to black boxes. But black boxes are placeholders for structural components or capacities thereof, to be identified by a complete mechanistic explanation of the capacities of the system. Thus, if psychological explanation is functional, as so many people assume, and psychological explanation is worthy of its name, then psychological explanation is mechanistic.

Once the structural aspects that are missing from a functional analysis are filled in, functional analysis turns into a more complete mechanistic explanation. By this process, functional analyses can be seamlessly integrated with mechanistic explanations, and psychology can be seamlessly integrated with neuroscience.

[20] This is consistent with the obvious point that in many contexts, it is useful to omit many details of a mechanistic explanation, whether functional or structural.

Most contemporary psychologists are unlikely to be surprised by our conclusion. A cursory look at current mainstream psychology shows that the boxological models that were mainstream in the 1970s and 1980s are becoming less popular. What is increasingly replacing them are structurally constrained models—mechanistic models. Much excitement surrounds methods such as neural imaging and biologically realistic neural network models, which allow psychologists to integrate their findings with those of neuroscientists. But some resistance to the integration of psychology and neuroscience remains, in both psychology and philosophy circles. We hope our argument will help soften such resistance.

For that purpose, it may also help to see the important role that autonomism about psychological explanation played in establishing cognitive psychology as a legitimate scientific enterprise. Autonomists hope to make room for an understanding of cognitive mechanisms that can proceed on its own, independently of neuroscience. Such a position was useful in the 1950's and 1960's, when knowledge of brain function was less advanced and we had only begun to think about how an integrated cognitive neuroscience would proceed. Before cognitive neuroscience could get off the ground, somebody had to characterize the cognitive phenomena for which neural explanations would be sought. The autonomist vision allowed experimental and theoretical psychologists to proceed with that task without having to wait for neuroscience to catch up. Now the discipline has advanced to the point that these pursuits can meaningfully come together, and there are tremendous potential benefits from affecting such integration.

Psychology should not content itself with the discovery of merely phenomenally adequate functional descriptions that fail to correspond to the structural components to be found in the brain. It should aim to discover mechanisms. To explain in cognitive psychology and neuroscience is to know the mechanisms, and explanation is what functional analysis has always been all about.

Bibliography

Aizawa, K. and C. Gillett (2009). "The (Multiple) Realization of Psychological and other Properties in the Sciences." Mind & Language **24**(2): 181-208.

Aizawa, K. and C. Gillett (2011). The Autonomy of Psychology in the Age of Neuroscience. Causality in the Sciences. P. M. Illari, F. Russo and J. Williamson. Oxford, Oxford University Press.

Aizawa, K. and C. Gillett (forthcoming). "Multiple Realization and Methodology in Neuroscience and Psychology." British Journal for the Philosophy of Science.

Bechtel, W. (2008). Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience. London, Routledge.

Bechtel, W. and J. Mundale (1999). "Multiple Realizability Revisited: Linking Cognitive and Neural States." Philosophy of Science **66**: 175-207.

Bechtel, W. and R. C. Richardson (1993). Discovering Complexity: Decomposition and Localization as Scientific Research Strategies. Princeton, Princeton University Press.

Bickle, J. (2003). Philosophy and Neuroscience: A Ruthlessly Reductive Approach. Dordrecht, Kluwer.

Bickle, J. (2006). "Reducing Mind to Molecular Pathways: Explicating the Reductionism Implicit in Current Cellular and Molecular Neuroscience." Synthese **151**: 411-434.

Block, N. (1997). "Anti-Reductionism Slaps Back." Mind, Causation, World, Philosophical Perspectives **11**: 107-133.

Block, N. and J. A. Fodor (1972). "What Psychological States Are Not." Philosophical Review **81**(2): 159-181.

Block, N. and G. Segal (1998). The Philosophy of Psychology. Philosophy 2: Further Through the Subject. A. C. Grayling. Oxford, Oxford University Press**:** 4-69.

Churchland, P. M. (1989). A Neurocomputational Perspective. Cambridge, MA, MIT Press.

Churchland, P. M. (2007). Neurophilosophy at Work. Cambridge, Cambridge University Press.

Churchland, P. S. (1986). Neurophilosophy. Cambridge, MA, MIT Press.

Cooper J. R., Bloom, F.E., and Roth, R.H. 1996. The Biochemical Basis of Neuropharmacology. Oxford: Oxford University Press.

Couch, M. (2005). "Functional Properties and Convergence in Biology." Philosophy of Science **72**: 1041-1051.

Crandall, B., G. Klein, et al. (2006). Working Minds: A Practitioner's Guide to Task Analysis. Cambridge, MA, MIT Press.

Craver, C. (2001). "Role Functions, Mechanisms, and Hierarchy." Philosophy of Science **68**(March 2001): 53-74.

Craver, C. F. (2004). “Dissociable Realization and Kind Splitting.” Philosophy of Science 71 (4): 960–971.

Craver, C. F. (2006). "When Mechanistic Models Explain." Synthese **153**(3): 355-376.

Craver, C. F. (2007). Explaining the Brain. Oxford, Oxford University Press.

Craver, C. and L. Darden (2001). Discovering Mechanisms in Neurobiology. Theory and Method in the Neurosciences. P. Machamer, R. Grush and P. McLaughlin. Pittsburgh, PA, University of Pittsburgh Press**:** 112-137.

Cummins, R. (1975). "Functional Analysis." Journal of Philosophy **72**(20): 741-765.

Cummins, R. (1983). The Nature of Psychological Explanation. Cambridge, MA, MIT Press.

Cummins, R. (2000). "How does it work?" vs. "What are the laws?" Two Conceptions of Psychological Explanation. Explanation and Cognition. K. F. C. and W. R. A. Cambridge, Cambridge University Press.

Dennett, D. C. (1978). Brainstorms. Cambridge, MA, MIT Press.

Edelman, G. M. and J. A. Gally (2001). "Degeneracy and Complexity in Biological Systems." Proceedings of the National Academy of the Sciences of the United States of America **98**(24): 13763-13768.

Figdor, C. (2010). "Neuroscience and the Multiple Realization of Cognitive Functions." Philosophy of Science **77**(3): 419-456.

Fodor, J. A. (1965). Explanations in Psychology. Philosophy in America. M. Black. London, Routledge and Kegan Paul.

Fodor, J. A. (1968a). Psychological Explanation. New York, Random House.

Fodor, J. A. (1968b). "The Appeal to Tacit Knowledge in Psychological Explanation." Journal of Philosophy **65**: 627-640.

Fodor, J. A. (1974). "Special Sciences." Synthese **28**: 77-115.

Fodor, J. A. (1997). "Special Sciences: Still Autonomous after All These Years." Philosophical Perspectives, 11, Mind, Causation, and the World: 149-163.

Gazzaniga, M., Ed. (2009). The Cognitive Neurosciences. Cambridge, MA, MIT Press.

Glennan, S. (2002). "Rethinking Mechanistic Explanation." Philosophy of Science **69**: S342-S353.

Glennan, S. (2005). "Modeling Mechanisms." Studies in History and Philosophy of Biological and Biomedical Sciences **36**(2): 443-464.

Gold, I. and D. Stoljar (1999). "A Neuron Doctrine in the Philosophy of Neuroscience." Behavioral and Brain Sciences **22**: 809-869.

Harman, G. (1988). Wide Functionalism. Cognition and Representation. S. Schiffer and S. Steele. Boulder, Westview**:** 11-20.

Heil, J. (2003). From an Ontological Point of View. Oxford, Clarendon Press.

Kalat, J. W. (2008). Biological Psychology. Belmont, CA, Wadsworth Publishing.

Kandel, E.R., Schwartz J.H., Jessell T.M. (2000). *Principles of Neural Science*, 4th ed. McGraw-Hill, New York.

Kaplan, D., and Craver, C. (forthcoming).

Keeley, B. (2000). "Shocking Lessons from Electric Fish: The Theory and Practice of Multiple Realizability." Philosophy of Science **67**: 444-465.

Kim, J. (1992). "Multiple Realization and the Metaphysics of Reduction." Philosophy and Phenomenological Research **52**: 1-26.

Kosslyn, S. M., W. I. Thompson, et al. (2006). The Case for Mental Imagery. New York, NY, Oxford University Press.

Machamer, P. K., L. Darden, and C. F. Craver (2000). "Thinking About Mechanisms." Philosophy of Science **67**: 1-25.

Marr, D. (1982). Vision. New York, Freeman.

Noppeney, U., K. J. Friston, and C. Price (2004). "Degenerate Neuronal Systems Sustaining Cognitive Functions." Journal of Anatomy **205**(6): 433-442.

O'Reilly, R. C. and Y. Munakata (2000). Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain. Cambridge, MA, MIT Press.

Piccinini, G. (2004). "Functionalism, Computationalism, and Mental States." Studies in the History and Philosophy of Science **35**(4): 811-833.

Piccinini, G. (2007). "Computing Mechanisms." Philosophy of Science **74**(4): 501-526.

Piccinini, G. (2008). "Computers." Pacific Philosophical Quarterly **89**(1): 32-73.

Piccinini, G. and A. Scarantino (2010). "Information Processing, Computation, and Cognition." Journal of Biological Physics.

Polger, T. W. (2004a). Natural Minds. Cambridge, MA, MIT Press.

Polger, T. (2009). "Evaluating the Evidence for Multiple Realization." Synthese **167**(3): 457-472.

Port, R. and T. van Gelder, Eds. (1995). Mind as Motion: Explorations in the Dynamics of Cognition. Cambridge, MA, MIT Press.

Posner, M. I. (2004). Cognitive Neuroscience of Attention. New York, NY, Guilford Press.

Price, C. J. and K. J. Friston (2002). "Degeneracy and Cognitive Anatomy." Trends in Cognitive Science **6**(10): 416-421.

Putnam, H. (1960). Minds and Machines. Dimensions of Mind: A Symposium. S. Hook. New York, Collier**:** 138-164.

Putnam, H. (1967a). The Mental Life of Some Machines. Intentionality, Minds, and Perception. H. Castañeda. Detroit, Wayne State University Press**:** 177-200.

Putnam, H. (1967b). Psychological Predicates. Art, Philosophy, and Religion. Pittsburgh, PA, University of Pittsburgh Press.

Putnam, H. (1975). Philosophy and our Mental Life. Mind, Language and Reality: Philosophical Papers, Volume 2. H. Putnam. Cambridge, Cambridge University Press**:** 291-303.

Shagrir, O. (1998). "Multiple Realization, Computation and the Taxonomy of Psychological States." Synthese **114**: 445-461.

Shagrir, O. (2006). "Why We View the Brain as a Computer." Synthese **153**(3): 393-416.

Shagrir, O. (2010). "Brains as Analog-Model Computers." Studies in the History and Philosophy of Science **41**(3): 271-279.

Shapiro, L. A. (2000). "Multiple Realizations." The Journal of Philosophy **XCVII**(12): 635-654.

Shapiro, L. A. (2004). The Mind Incarnate. Cambridge, MA, MIT Press.

Shepherd, G.M. 1994. *Neurobiology*. 3rd edition. London: Oxford University Press.

Sporns, O., Tononi, G., and R. Kötter (2005). “The Human Connectome: A Structural Description of the Human Brain.” PLoS Computational Biology, **1**(4): 245-250.

Stich, S. (1983). From Folk Psychology to Cognitive Science. Cambridge, MA, MIT Press.

Turing, A. M. (1950). "Computing Machinery and Intelligence." Mind **59**: 433-460.