Synthese
DOI 10.1007/s11229-015-0783-4

CrossMark

S.I. : NEUROSCIENCE AND ITS PHILOSOPHY

# The cognitive neuroscience revolution

**Worth Boone**[1] · **Gualtiero Piccinini**[2]

Received: 22 September 2014 / Accepted: 22 May 2015


**Abstract** We outline a framework of multilevel neurocognitive mechanisms that incorporates representation and computation. We argue that paradigmatic explanations in cognitive neuroscience fit this framework and thus that cognitive neuroscience constitutes a revolutionary break from traditional cognitive science. Whereas traditional cognitive scientific explanations were supposed to be distinct and autonomous from mechanistic explanations, neurocognitive explanations aim to be mechanistic through and through. Neurocognitive explanations aim to integrate computational and representational functions and structures across multiple levels of organization in order to explain cognition. To a large extent, practicing cognitive neuroscientists have already accepted this shift, but philosophical theory has not fully acknowledged and appreciated its significance. As a result, the explanatory framework underlying cognitive neuroscience has remained largely implicit. We explicate this framework and demonstrate its contrast with previous approaches.

**Keywords** Cognitive neuroscience · Multilevel mechanisms · Explanation · Integration · Computation · Representation

The authors are listed alphabetically—the paper was thoroughly collaborative. Otávio Bueno graciously arranged for the double blind refereeing of this paper; thanks to him and the anonymous referees for helpful comments. Thanks to our audiences at Georgia State University, Washington University in St. Louis, the 2014 Society for Philosophy and Psychology meeting, 2014 Central APA meeting and to our APA commentator, Robert Rupert. Thanks to Sergio Barberis, Mazviita Chirimuuta, and Corey Maley for helpful comments. Thanks to Elliott Risch for editorial assistance. This material is based on work supported in part by a University of Missouri research award to Gualtiero Piccinini.

✉ Gualtiero Piccinini
piccininig@umsl.edu

1 University of Pittsburgh, Pittsburgh, USA

2 University of Missouri – St. Louis, St. Louis, MO, USA



Journal: 11229-SYNT Article No.: 0783 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

# 1 Introduction

The traditional framework of cognitive science included (aspects of) six disciplines: psychology, computer science, linguistics, anthropology, neuroscience, and philosophy. These six disciplines were supposed to work together towards understanding cognition in accordance with a neat division of labor, to which many practitioners conformed. On one side stood psychology, with the help of computer science, linguistics, anthropology, and philosophy; on the other side stood neuroscience. Psychology etc. studied the functional or cognitive level, or—in Marr's terminology—the computational and algorithmic levels; neuroscience investigated the neural, mechanistic, or implementation level. Explanations at these two levels were considered distinct and autonomous from one another.

This division of labor leaves no room for cognitive *neuro*science. Indeed, from this perspective, the very term "cognitive neuroscience" is almost an oxymoron, because neuroscience is supposed to deal with the mechanisms that implement cognitive processes, not with cognition proper. Yet cognitive neuroscience has emerged as the new mainstream approach to studying cognition. What gives?

In this paper, we argue that cognitive science as traditionally conceived is on its way out and is being replaced by cognitive neuroscience, broadly construed. Cognitive neuroscience is still an interdisciplinary investigation of cognition. It still includes (aspects of) the same six disciplines (psychology, computer science, linguistics, anthropology, neuroscience, and philosophy). But the old division of labor is gone, because the strong autonomy assumption that supported it has proven wrong.

The scientific practices based on the old two-level view (functional/cognitive/computational vs. neural/mechanistic/implementation) are being replaced by scientific practices based on the view that there are *many* levels of mechanistic organization. No one level has a monopoly on cognition proper. Instead, different levels are more or less cognitive depending on their specific properties. The different levels and the disciplines that study them are not autonomous from one another. Instead, the different disciplines contribute to the common enterprise of constructing multilevel mechanistic explanations of cognitive phenomena. In other words, there is no longer any meaningful distinction between cognitive psychology and the relevant portions of neuroscience—they are merging to form cognitive neuroscience. Or so we will argue.

By contrast, many philosophers still insist that psychological explanation is distinct and autonomous from neuroscientific explanation. Some argue that psychological explanations can be satisfactory without being mechanistic (e.g., Weiskopf 2011, but see Povich forthcoming for a rejoinder). Others argue that representational and computational explanations of cognition belong in an autonomous psychology not in neuroscience (Fodor 1998; Burge 2010). A somewhat independent view, which also stands in contrast to our framework, is that computational explanation is not mechanistic (Rusanen and Lappi 2007; Shagrir 2010a; Chirimuuta 2014). In addition, there are scientists who argue that current neuroscience is wrong-headed and should be reformulated in light of a rigorous computational psychology (Gallistel and King 2009). While the latter view may be seen as consistent with our integrationist framework, in our opinion it underestimates the extent to which current neuroscience is empirically well grounded and should constrain our cognitive explanations.





Journal: 11229-SYNT Article No.: 0783 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

We have two primary, closely related goals. The first is to explicate the explanatory framework underlying contemporary cognitive neuroscience, contrasting it with traditional cognitive scientific explanation. The second is to soften current resistance to the mechanistic integration of psychology and neuroscience. We proceed as follows. After reconstructing the received view of explanation in cognitive science (Sect. 2), we briefly indicate why traditional responses to the received view fail to square with cognitive neuroscience as we understand it (Sect. 3). We then articulate a framework of multilevel neurocognitive mechanisms (Sect. 4) and the levels that constitute them (Sect. 5). We conclude by highlighting three important aspects of cognitive neuroscience that illustrate our framework: the incorporation of experimental protocols from cognitive psychology into neuroscience experiments, the development and evolution of functional neuroimaging, and the movement toward biological realism in computational modeling (Sect. 6). One important consequence of the picture we advance is that neither structures nor functions have primacy in individuating the scientific kinds of cognitive neuroscience. The upshot is that explanation in cognitive neuroscience is multilevel, mechanistic, computational, and representational.

## 2 Cognitive science as traditionally conceived

The *cognitive revolution* of the 1950s is most often juxtaposed against the behaviorist program it supplanted. By contrast with behaviorism's methodology and metaphysics, which is widely assumed to reject the postulation of cognitive states and processes, cognitive science explicitly postulates internal cognitive states and processes to explain intelligent capacities. An important motivation for this approach came from the analogy between cognitive systems and digital computers. Computers possess internal states and processes that contribute to their capacities, some of which—playing chess, solving problems, etc.—are capacities that require intelligence in humans. Since it's patently legitimate to explain a computer's capacities in terms of its internal states and processes, cognitive scientists argued that it is equally legitimate to explain human cognition in terms of internal states and processes. More importantly, the internal states and processes of computers are representations and computations, which are typically considered cognitive notions. Thus, the argument continues, it is legitimate to explain human cognition in terms of computations and representations. Indeed, in this tradition cognition is often *identified* with some form of computation—more specifically, some form of *digital* computation over representations (e.g., Newell and Simon 1976; Anderson 1983; Johnson-Laird 1983; Pylyshyn 1984).

This focus on the contrast between behaviorism and cognitive psychology often obscures some of the substantive commitments that came out of the cognitive revolution. At all stages of Western history, available technology has constrained the analogies used to think about the operations of the human mind and body. For instance, water technologies—pumps, fountains, etc.—provided the dominant metaphor behind the ancient Greek concept of the soul—the 'pneuma'—and the humorist theories that dominated Western medicine for 2000 years (Vartanian 1973); the gears and springs of clocks and wristwatches played a similar role for early mechanist thinking during the enlightenment (e.g., La Mettrie's *L'Homme Machine*, 1748); hydraulics for Freud's



Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

concept of libido; telephone switchboards for behaviorist theories of reflexes; and so on.[1] It is no coincidence that the cognitive revolution co-occurred with the advent of computers.

Whenever technology guides thinking about the human mind or body, there is risk that the analogy is taken too far. While it may be true that cognition involves transitions between internal states analogous to computations of some kind, the commitments of traditional cognitive science go far beyond this basic point. Specifically, the analogy between cognition and computation has been taken to imply that cognition may be studied *independently* of the nervous system. The main rationale for this autonomy is that digital computers—more specifically, universal, program-controlled digital computers—reuse the same hardware for the different programs (i.e., software) they execute. Each particular program explains a specific capacity, while the hardware remains the same. By the same token, a widespread assumption of traditional cognitive science is that the brain is a universal, program-controlled digital computer; therefore, cognition can be studied simply by figuring out what programs run on such a computer, without worrying over the details of the wetware implementation of those programs (e.g., Fodor 1968b; Newell and Simon 1976; Pylyshyn 1984). Additionally, many who thought of the brain simply as some kind of digital computer (without assuming that it is universal and program-controlled) nonetheless agreed that cognition could be explained independently of neuroscience (e.g., Cummins 1983).[2]

A close ally of this computer analogy and its rationale for the autonomy of psychology is the view that psychological explanation is different in kind from neuroscientific explanation. According to this view, psychological explanation captures cognitive functions and functional relations between cognitive states and capacities, whereas neuroscientific explanation aims at the structures that implement cognitive functions. The two types of explanation are supposed to place few constraints on one another with the upshot that each can proceed independently from the other.

The resulting picture of cognitive science is that psychology studies cognition in functional terms, which are autonomous from the non-cognitive mechanisms studied by neuroscience. Aspects of this two-level picture can be found in the writings of many philosophers of cognitive science. Here are a few stark examples:

> 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, p. 149).

> [I]n the language of neurology …, presumably, notions like computational state and representation aren't accessible (Fodor 1998, p. 96).

[1] See Daugman (1990) for more detailed discussion of the role of technology and metaphor in the study of the human mind and body.

[2] A computer is universal just in case it can compute any computable function until it runs out of memory and time. A computer is program-controlled just in case it computes different functions depending on which program it executes. Contemporary digital computers are both universal and program-controlled. Different kinds of analogies may be drawn between digital computers and brains, some of which are stronger than others (cf. Piccinini 2008, Sect. 5 for a more detailed discussion). At the same time, it was widely recognized that there are significant architectural and performance differences between artificial digital computers and natural cognitive systems.



Journal: 11229-SYNT Article No.: 0783 TYPESET ☑ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

> We could be made of Swiss cheese and it wouldn't matter (Putnam 1975, p. 291).

> [F]unctional analysis [which includes psychological explanation] puts very indirect constraints on componential analysis [i.e., mechanistic explanation] (Cummins 1983, p. 29; 2000, p. 126).

> [E]xplanation in terms of distinctively psychological representational notions is, as far as we now know, basic and ineliminable (Burge 2010, p. 27).

These philosophers were, and in some cases still are, trying to capture what cognitive scientists were doing at the time. And while cognitive scientists were perhaps less explicit about the two-level picture, something similar to this view can be found in many landmark works that came out during the heyday of classical cognitive science (e.g., Newell and Simon 1976; Newell 1980; Marr 1982; Anderson 1983; Johnson-Laird 1983; Pylyshyn 1984).

## 3 Traditional responses to cognitive science

This traditional two-level picture of cognitive science fails to capture explanation in contemporary cognitive *neuro*science. Cognitive neuroscience strives to explain cognition on the basis of neural mechanisms and thus involves integration, not autonomy, between psychology and neuroscience. After the cognitive revolution, the mechanistic integration of psychology and neuroscience amounts to another paradigm shift: the cognitive *neuro*science revolution. In later sections we will argue that this new revolution requires a different way of thinking about levels, cognitive explanation, representation, and computation. The resulting explanatory framework, multilevel neurocognitive mechanisms, is what we aim to articulate in this paper.

In seeking an account of explanation in cognitive neuroscience, let's begin with two traditional responses to the two-level picture—reduction and elimination. While we lack the space for detailed treatment, we briefly argue that these traditional responses to cognitive science fail to adequately capture the kind of integration found in cognitive neuroscience. These arguments will motivate our positive proposal (Sect. 4).

One traditional alternative to autonomy is to eliminate the theoretical constructs posited by psychology *in favor of* the theoretical constructs posited by neuroscience. The model for such eliminativism is the past elimination of theoretical constructs, such as epicycles, phlogiston, or the ether, from past scientific theories. Just as those theoretical posits were eventually eliminated from our scientific theories of, respectively, planetary motion, heat, and the transmission of radiation through space, so the theoretical posits of psychology, such as the language-like mental representations posited by classical cognitive psychologists, should be eliminated in favor of posits that are more amenable to neuroscience (Churchland 1981, 1986).

If eliminativism is construed radically enough—that is, as the literal elimination of any science of cognition other than neuroscience—it offers a partial solution to the problem at hand. That problem is to understand how the disciplines that study cognition fit together and how cognition ought to be explained. If any discipline other than neuroscience is eliminated, the first half of the problem is solved: since the other





Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

disciplines no longer exist, we don't need to worry about how they fit together with neuroscience. But this radical construal is hardly a solution to the most interesting part of the problem—how to explain cognition.

Contemporary cognitive neuroscience aims to explain cognition on the basis of neural computation over neural representations (more on this below). If the eliminativist approach implies that cognition itself—and all "cognitive" theoretical posits, such as representation, computation, or information processing—should be eliminated or at least deflated (cf. Ramsey 2007), then we are faced with a solution that is antithetical to cognitive neuroscience.

Another alternative to traditional (two-level) cognitive science is to *reduce* psychological theoretical posits to neuroscientific theoretical posits. The models for this reductionist strategy come from examples from physics, such as the reduction of classical thermodynamics to statistical mechanics or the reduction of Newton's theory of gravitation to a special case of Einstein's theory of General Relativity. The main difficulty for this reductionist approach in cognitive neuroscience is that, even assuming that it works for some physical theories (which has been debated), psychological and neuroscientific explanations lack the appropriately general mathematical formulation to be conducive to such reductions (cf. Cummins 2000).

Nevertheless, some have argued that when we can intervene on molecular structures in the brain and affect some cognitive behavior, the specific molecular events "directly explain" the behavioral data, we thereby reduce the relevant cognitive capacity to the relevant molecular events, and we thereby obviate the need for intermediate levels of explanation (cf. Bickle 2003, 2006). The main problem with this form of reductionism is that specific molecular events are at best only *partial* explanations of cognitive phenomena. It is one thing to correlate specific molecular events with cognitive phenomena via some specific intervention; to actually explain a cognitive phenomenon on the basis of molecular events requires determining the ways in which the molecular events are causally relevant to the production of the phenomenon of interest. Molecular events are only relevant to the extent that they occur within specific neural structures, and locating the relevant neural structures requires more than purely molecular neuroscience. In addition, even identifying a molecular event within a neural structure that contributes to a cognitive behavior falls short of a full explanation. A full explanation requires identifying how molecular events contribute to relevant neural events, how relevant neural events contribute to circuit and network events, how those in turn contribute to relevant systems-level events, and finally how the relevant systems, appropriately coupled with the organism's body and environment, produce the behavior. These intermediate links in the causal-mechanistic chain are crucial to connecting molecular events to cognitive phenomena in a way that is explanatory, as opposed to merely correlational. And identifying these intermediate level structures and their causal contributions requires going well beyond molecular neuroscience (cf. Craver 2007; Bechtel 2008).

In spite of their respective limitations, both eliminativism and reductionism put pressure on the received view of cognitive science—most helpfully, by pointing out that cognitive scientists who ignore neuroscience do so at their peril and by pushing towards the integration of psychology and neuroscience. But neither eliminativism nor reductionism offers a satisfactory framework for explanation in *cognitive* neuro-



Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

science: the former insofar as it neglects cognition altogether; the latter because it offers only partial explanations that lack the necessary contextual factors provided by intermediate levels of analysis.

## 4 Multilevel neurocognitive mechanisms

Cognitive neuroscience stands in stark contrast to the traditional two-level picture of cognitive science. Broadly, cognitive neuroscience is the scientific study of how neural activity explains cognition and the behavior it gives rise to. Cognitive neuroscientists study nervous systems using many techniques at many levels. They study how cortical areas and other neural systems contribute to various cognitive capacities, how the capacities of those systems are explained by the operations of the neural subsystems that compose them (columns, nuclei), how networks and circuits contribute to their containing systems, how neurons contribute to networks and circuits, and how sub-neuronal structures contribute to neuronal capacities. Analyzing systems across such varied levels involves coordinating techniques ranging from molecular neuroscience and genetics to neurophysiology, neuroimaging, mathematical analysis, computational modeling, and a wide range of behavioral tasks.

Cognitive neuroscience thus strives to explain cognitive phenomena by appealing to and analyzing (both separately and conjointly) multiple levels of organization within neural systems. *Multilevel mechanisms* have recently been proposed as a framework for thinking about the relations between these levels of organization. A multilevel mechanism is a system of component parts and wholes in which the organized capacities of the component parts constitute (and thus mechanistically explain) the capacities of the whole (e.g., Craver 2007). Some mechanists prefer to define mechanisms in terms of operations, activities, or interactions rather than capacities (e.g., Glennan 2002; Bechtel 2008). We see these different formulations as equivalent for the purposes at hand because operations, activities, and interactions can be seen as manifestations of capacities (Piccinini unpublished). Note that it may take multiple capacities organized in specific ways to bring about specific operations, activities, or interactions. In this section, we expand this framework, arguing for a specific understanding of cognitive neuroscience as a science directed at integrated multilevel neurocognitive mechanisms.[3]

Multilevel neurocognitive mechanisms have an iterative structure: at any level, each component of the mechanism is in turn another mechanism whose capacities are explained by the organized capacities of *its* components; and each whole mechanism is itself a component part that contributes to the capacities of a larger whole. This multilevel iterative structure tops off in the capacities of whole organisms and their interactions with other organisms, which are studied by social neuroscience and neuroeconomics; it bottoms out in structures—such as the atoms that compose

[3] Some argue that at least some explanations in cognitive neuroscience are not mechanistic but are instead "dynamical" (e.g., Chemero and Silberstein 2008). We lack the space to discuss this putative alternative to mechanistic explanation, except to point out that mechanistic explanations are often dynamical in the relevant sense (cf. Bechtel and Abrahamsen 2013) and thus are consistent with describing the dynamics of a system, whereas dynamical descriptions may or may not be explanatory in the relevant sense (cf. Kaplan and Craver 2011).





Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

neurotransmitters—that fall outside the disciplinary boundaries of cognitive neuroscience.

Cognitive neuroscience is not the only science that explains mechanistically, but it is one of the few whose mechanisms perform computations over representations (cf. Bechtel 2008, 2015). There is a large literature on what constitutes computation and representation and we cannot do justice to these topics here. For present purposes, it will suffice to sketch an account of computation and representation that squares with the framework of multilevel neurocognitive mechanisms.

A vehicle carries *semantic information* about a source just in case it reliably correlates with the states of the source (Dretske 1981; Piccinini and Scarantino 2011; Scarantino 2015). For instance, the spike trains generated by neurons in cortical area V1 reliably correlate with the presence and location of edges in the visual environment; thus, they carry semantic information about the presence and location of edges in the visual environment. But correlation alone is insufficient for representation.

A vehicle *represents* a source just in case it has the function of carrying information about the source (Dretske 1988; Morgan 2014). For a vehicle to have such a function, the information it carries must be used by some part of the system in which it is embedded. The information is used by the system to the extent that it's causally relevant to other operations of the system. In our example, the spike trains generated by neurons in V1 have the function of carrying information about the visual environment because this information is used by downstream areas for further processing of visual stimuli—i.e., it is causally relevant to the operations of those downstream areas. Thus, in the relevant sense, V1 neurons represent the presence and locations of the edges with which they correlate.

Finally, a system performs *computations* just in case it manipulates vehicles in accordance with rules that are sensitive to inputs and internal states and are defined in terms of differences between different portions of the vehicles it manipulates. Which computations are performed by a system depends on its specific mechanistic properties—its component types, its vehicle type, and the rules it follows. That is, computation here is defined non-semantically based on the mechanistic properties of the system and the vehicles it manipulates. Although computation can occur in the absence of representation, processing representations is a form of mechanistic computation (Piccinini and Scarantino 2011; cf. Fresco 2014; Milkowski 2013).

A distinctive feature of neural systems is that they pick up information from the environment and organism, transmit it through the system via appropriate signals (neural representations), and process such signals in conjunction with pre-existing representations and rules of manipulation (neural computation) in order to generate further signals that regulate the organism's behavior. This appeal to representation and computation distinguishes mechanistic explanations in cognitive neuroscience from mechanistic explanations in many other sciences.

The above account of computation is diametrically opposed to persistent views of computation that draw a stark contrast between computational and mechanistic explanations.[4] Such views maintain that computations are *abstract* or *mathematical* in a

[4] A recent example: "My key claim is that the use of the term 'normalization' in neuroscience retains much of its original mathematical-engineering sense. It indicates a mathematical operation—a computation—not





Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

way that evades mechanistic explanation. While it's true that computation can be mathematically characterized, however, the physical computations performed by nervous systems (and artificial computers, for that matter) are functions performed by concrete mechanisms.[5] Like other functions, information processing via neural computation is performed by mechanisms—specifically, *neurocomputational* mechanisms. With this said, an important caveat is that computing mechanisms, like all mechanisms, can be characterized at different levels of abstraction. This is an integral aspect of multilevel mechanistic explanation, though one that has been a source of recent controversy.

The mechanistic framework has recently been construed as a call for maximal detail in explanation and a rejection of abstraction. A number of recent criticisms have been developed along these lines (Barberis 2013; Barrett 2014; Levy and Bechtel 2013; Levy 2013; Chirimuuta 2014; Weiskopf 2011). A common idea behind these objections is that the multilevel mechanistic framework is committed to the claim that the explanatory power of a model is primarily a function of the amount of detail contained in its description of a particular mechanism—viz. the more detail, the better the explanation. Thus, according to this interpretation, mechanistic integration eschews any valuable role for abstraction in explanation.[6]

As these critics point out, many forms of explanation in cognitive and systems neuroscience proceed through systematic abstraction away from the particular details of a target system. This is, for instance, how neuroscientists come to characterize something like lateral inhibition as a general type of organization of neural circuitry found in different brain regions—e.g., peripheral somatosensory and visual processing both exhibit this kind of organization. The details—the particular kind of excitatory cell, inhibitory interneuron, number and strength of synapse, etc.—are often irrelevant to understanding why lateral inhibition is a useful form of circuitry and why it crops up in so many circumstances in which these details do in fact differ. It is thus easy to see why a view in which explanatory power is tied solely to detail of description would face serious problems in cognitive neuroscience.

But it would be a mistake to conclude that when an explanation intentionally excludes some details, the explanation is thereby rendered non-mechanistic. To

Footnote 4 continued
a biological mechanism" (Chirimuuta 2014, p. 124). Chirimuuta also cites some neuroscientists who draw a similar contrast between computations and mechanisms.

[5] Not all *mathematical* models in cognitive neuroscience ascribe computations to the nervous system; only those that explain phenomena through computations performed by the target systems do so.

[6] In fairness to the critics, some mechanists may give the impression of advocating such a view: "the more accurate and detailed the model is for a target system or phenomenon the better it explains that phenomenon, all other things being equal" (Kaplan 2011, p. 347). Kaplan points out that some details may be omitted from a model either for reasons of computational tractability or because they are unknown. Similarly, Craver writes: "Between sketches and complete descriptions lies a continuum of mechanism schemata whose working is only partially understood" (Craver 2007, p. 114). To drive this point home, Craver aligns the sketch-schema-mechanism axis with the epistemic axis of "how possibly-plausibly-actually": "Progress in building mechanistic explanations involves movement along both the possibly-plausibly-actually axis and along the sketch-schema-mechanism axis" (Craver 2007, p. 114). Contrary to what Craver appears to imply, progress may consist in abstracting away from irrelevant details to construct an appropriate schema, and in some epistemic contexts even a mechanism sketch may provide all the explanatory information that is needed (more on this in this section). And in fairness to Craver and Kaplan, we should note that there are also passages where they accept that abstraction and idealization play legitimate roles in explanation.



Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

the contrary, proponents of the mechanistic framework have often pointed out that *abstracting away* from irrelevant details is as important to mechanistic explanation as *including* relevant details (e.g., Piccinini and Craver 2011; see Boone and Piccinini unpublished for a more detailed treatment). Which details ought to be included and excluded depends on various features of explanatory context. The concepts of mechanism *sketches* and *schemata* were designed to capture this aspect of mechanistic explanations (Machamer et al. 2000). Mechanism sketches involve omissions of as yet unknown details; mechanism schemata involve deliberate omissions of detail, capturing the bare relevant causal structure of a system.

Examples of schemata abound in neuroscience. A much-discussed example, which is particularly relevant to the present context, is the Hodgkin-Huxley model of the action potential (Hodgkin and Huxley 1952). The Hodgkin-Huxley model explains the voltage profile of the action potential in terms of a neural membrane's changing voltage conductivity. Lower-level mechanistic details about how changes in membrane permeability arise were omitted from the model, initially because they were unknown (Hodgkin and Huxley 1952, p. 541), but also later because this omission affords the model greater generality (Schaffner 2008; Levy 2013; Chirimuuta 2014, p. 141). The Hodgkin-Huxley model has been described as non-explanatory (Bogen 2005), as providing a non-mechanistic explanation (Weber 2005, 2008), and as a mere sketch because it omits information about the role of ion channels in allowing membrane permeability (Craver 2007). In our view, none of these characterizations fully hit the mark. Rather, the HH model is an example of a mechanism sketch that evolved into a mechanism schema—it explains a phenomenon (the action potential) at one mechanistic level (changes in membrane conductivity) while abstracting away from lower mechanistic levels (ion channels and their components).

As the preceding example illustrates, mechanistic explanations—particularly those that involve computations and representations in the sense outlined above—are often presented in the form of (interpreted) mathematical or computational models. Typically, such models become analytically insoluble and computationally intractable if they include too much detail about their target systems. As such, issues relating to solubility and tractability provide another motivation for the exclusion of detail from models of neurocomputational mechanisms. Issues regarding tractability are ubiquitous in computational neuroscience given the vast array of biological detail that could potentially figure into modeling scenarios.

For instance, one controversial but extremely common assumption among computational neuroscientists is that individual neurons can be treated as integrating a linear sum of dendritic inputs, and then initiating an action potential when that sum reaches a threshold. The dynamics of actual neurons are more complex than this model suggests, which in turn has led to the development of more complex models—e.g. Waxman (1972) provides an alternative model, which introduces nonlinearities into the branching regions of the dendritic (input) and axonal (output) trees, rather than treating those regions as, respectively, collecting and distributing charge linearly. But the basic, linear treatment of dendritic input integration has been a powerful tool in a wide variety of modeling contexts. One explanation for the success of these simplified modeling strategies is that they capture important aspects of neural responses that are adequate for particular epistemic purposes.





Journal: 11229-SYNT Article No.: 0783 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

While this topic raises a number of interesting issues, one clear takeaway is that an important skill of mathematical and computational modelers is to capture all and only those features of the systems they study that are needed to explain phenomena of interest, often by introducing appropriate idealizations and simplifications. Those idealizations and simplifications allow modelers to represent systems to desired degrees of approximation while maintaining mathematical solubility or computational tractability (cf. Humphreys 2004; Piccinini 2007; Winsberg 2010; Weisberg 2013).[7] Explanations of computational or information processing mechanisms often require these forms of detail omission. What is crucial to appreciate here is that computation and information processing do *not* lead outside the multilevel mechanistic framework but are instead best seen as a special case of it.[8]

Relatedly, as a matter of methodology, we are often interested in one aspect (some components or capacities) of a mechanism at the expense of other aspects (other components or capacities). This is one type of mechanism *sketch*, or partial (elliptical) mechanistic explanation. Consider what it takes to explain why a mechanism functions differently than it normally does. Explaining a deviation from normal functioning may require simply pointing out what's different in the relevant case, while omitting the rest of the mechanism (Van Eck and Weber 2014). For instance, to explain why certain patients have left-side hemineglect (roughly: inattention to and unawareness of the left side of visual space) it may be enough to point out that such patients suffered damage to the contralateral (right-side) cortical areas responsible for spatial attention, without providing details about the mechanisms involved in normal spatial attention and consciousness.

Special cases of this type of mechanism sketch are descriptions of computational (the function computed and why it is adequate to the task, cf. Shagrir 2010b) and algorithmic (the computational operations and representations) levels of a system, which omit details about the components that carry out the algorithm. There is certainly value to such approaches in cognitive neuroscience, particularly in the context of discovery. Marr (1982) argued that a neural "algorithm" is "likely to be understood more readily by understanding the nature of the problem being solved than by examining the mechanism (and the hardware) in which it is embodied" (Marr 1982, p. 27). Marr's arguments have often been cited in defense of autonomist views of cognitive

[7] Issues related to tractability and solubility of mathematical models quickly get into deeper philosophical water than can be adequately treated here. Such issues spread across most domains of scientific inquiry. For instance, foundational work in continuum mechanics—i.e. the Navier–Stokes equations—developed around failures to model the flows of fluids through containers as trajectories of point particles; rather, the Navier–Stokes equations describe velocity fields at given points in space and time (see Batterman 2013 for an extended discussion). The extent to which the successes of these "top-down" modeling strategies can be treated merely as idealizations and approximations rather than reflecting more fundamental differences in the phenomena under investigation and our understanding of those phenomena at different levels of analysis is currently a topic of rich philosophical debate.

[8] This is not to say that all analyses of neural computation or information-processing are mechanistic. Some focus only on the information content and efficiency of a neural code without saying anything about the processing mechanisms (Dayan and Abbott 2001, xiii; Chirimuuta 2014, p. 143ff). These models are not especially relevant here because they do not provide the kind of constitutive explanations that are the present topic, and that functional analysis and mechanistic explanation are competing accounts of.





Journal: 11229-SYNT Article No.: 0783 TYPESET ☑ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

science.[9] We have a different take, consistent with seeing computational and algorithmic accounts (in Marr's sense) as mechanism sketches (or schemata to the extent that underlying details are known but deliberately omitted).

Understanding the capacities of a system often requires looking "up" to situate the system within some higher-level mechanism or environmental context as much as looking "down" to understand how those capacities are implemented by the lower level components, their capacities, and organization. In addition, more may be known about the mechanistic or environmental context of a system than its components and their operations. In such cases, investigators may be forced to constrain their models primarily by examining the problem being solved rather than the components and their operations, even though the likely result of such a method is a "how-possibly" model that falls short of explaining how the system *actually* works. Much of Marr's work belongs in this how-possibly category. We certainly face some of the same problems in contemporary cognitive neuroscience, but the field has developed to the point where integration, rather than autonomy, is the appropriate framework. The computational-level descriptions Marr and others sought are best construed as a valuable step along the way to integrated multilevel mechanistic explanations. It is no longer enough to simply home in on ways in which problems *might* be solved in the brain; contemporary cognitive neuroscience aims to understand how those problems are *actually* solved in the brain.

## 5 Neurocognitive levels

A primary motivation behind the traditional autonomist picture of cognitive science is the idea that functions can be understood independently from the structures that perform them. Therein lies the putative distinction between the "functional" level, which is cognitive, representational, and computational, and the "structural" level, which is non-cognitive, mechanistic, and implementational. Our account of multilevel neurocognitive mechanisms adopts a different notion of neurocognitive levels, which undermines this traditional dichotomy.

Contrary to the received view, there is no single "functional," "cognitive," or "representational/computational" level of explanation, standing in opposition to a single (or even multiple) "structural," "neural," or "implementational" level(s).[10] In this section we analyze each of these concepts from the perspective of neurocognitive mechanisms in order to highlight how our integrationist framework improves upon traditional autonomist and reductionist views.

In the first place, every level of a multilevel mechanism is both *functional* and *structural*, because every level contains structures performing functions. This stands in stark contrast to traditional views that maintain that structural analyses and functional analyses are distinct and autonomous from one another. Traditional reductionists—

[9] Bechtel and Shagrir (forthcoming) is a good entry into the extensive literature on Marr's levels, including how they might fit within a mechanistic framework. We cannot do justice to that debate here.

[10] This point is reminiscent of Lycan's underappreciated critique of "two-levelism" (Lycan 1990). But Lycan lacked the accounts of mechanistic explanation and computational explanation that have been developed in detail in the past decade, and that provide the foundation that we are building upon.



Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

e.g. type physicalists (Smart 1959)—strove to identify mental types with physical types. As a result, they may be interpreted as focusing on structural properties at the expense of functional properties, relegating the latter to "second order states" of physical types (Smart 2007). Traditional functionalists do the opposite: they give primacy to functional properties at the expense of structural properties (e.g., Putnam 1967; Fodor 1968a). This is a somewhat unorthodox way of characterizing these views, so some brief unpacking is in order.

Classical reductionist views of the mind-brain relation, specifically type identity theorists, look to identify higher-order kinds (e.g. mental kinds, like "pain") with corresponding physical kinds. These reductive views were developed in contrast to dualism: the view that the mind and brain are distinct kinds of substance. Dualists have a notoriously difficult time specifying the means by which these distinct substances interact; type identity theorists provide a dissolution of this problem. To say that water is $H_2O$ just is to identify a higher-level kind with a lower-level physical kind—an arrangement of atoms. With such an identity in hand, it is illegitimate to wonder how water and $H_2O$ interact. In a similar vein, type identity theorists argued that mental kinds, like pain, could be identified with particular neural kinds, like C-fibers firing. This identification dissolves dualistic concerns about how mental states interact with bodily states. What is noteworthy for present purposes is that the defining features of the kinds that figure into higher-level analyses just are the lower-level physical features common to instances of those kinds. This identification with lower-level physical features downplays the role of functional features of those higher-level kinds. Water is not individuated by its ability to quench thirst, nourish plants, etc., nor pain with its role in avoidance of noxious stimuli, protecting injured body parts, etc. Instead both kinds are identified with particular physical types, which possess the relevant causal powers that are, incidentally, associated with these functions.

Classic functionalist views turn this story on its head: the defining features of higher-level kinds, according to such views, are their functional features while the structural features are incidental. The crux of this disagreement between functionalists and reductionists turns on multiple realizability—i.e., the claim that the same function can be realized in distinct physical substrates. Carburetors provide a classic example: they are defined by their function in internal combustion engines (mixing fuel and air); they retain this function over variations in the stuff they're made of (e.g. cast iron, zinc, aluminum, plastic) and the types of engine they're found in (e.g. car, motorcycle, lawn mower). Putnam's original arguments for autonomy in the 1960s were based on this insight (e.g. Putnam 1967). Fodor took up the torch and used multiple realizability to argue for the general autonomy of the special sciences from lower-level sciences—physics, in particular (Fodor 1968a, b, 1974). The idea behind these arguments is that, while higher-level states are token identical to particular lower-level physical states, there is no single lower-level physical kind for the higher-level states to be identified with. Rather, when higher-level kinds are realized, the underlying physical kinds will form unruly disjunctions (e.g. cast iron OR zinc OR aluminum OR plastic); the only thing tying this disjoint set of physical features together is the higher-level kind itself (e.g. the function, "mixing fuel and air"). Thus nothing is added to the higher-level analysis by looking at its realizers.



Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

Neither of these approaches adequately captures the main thrust of work in cognitive neuroscience, because that work is aimed at understanding the complex interplay *between* structure and function. By contrast, the multilevel mechanistic framework we are advocating adequately captures this aspect of cognitive neuroscientific explanations; in our framework, functions constrain structures and vice versa. Functions *cum* contextual factors—i.e. mechanistic context—constrain the range of structures that can perform those functions. Similarly, structures *cum* contextual factors determine the range of functions those structures can perform. In this framework, neither structures nor functions are given primacy over the other; neither can explain cognition without the other.

Any given structure is only capable of performing a restricted range of functions. For an everyday example, consider again the functions that can be associated with water. Structural facts about the chemical composition of water both enable and restrict its ability to perform certain functions. For instance, the facts that water is liquid at physiological temperatures and is composed of hydrogen (positively charged) and oxygen (negatively charged) make it appropriate for dissolving ionic compounds into ions essential for normal cell function. Contextual factors—like ambient temperature and available compounds—combine with structural factors to determine the appropriate range of functions. In the context of cognitive science, similar observations abound. Consider for instance that neurons have a refractory period, during which they cannot fire. This refractory period restricts a neuron's maximum firing rate to about 1000 Hz, which in turn limits the kind of codes by which the brain can encode and transmit information. The structural properties that determine the recovery period of a neuron—blocks that prohibit influx of sodium ions through voltage-gated channels—limit the encoding and signaling functions neurons can perform.

In the other direction, any function can only be performed by a restricted range of structures. For an everyday example, reconsider the example of a carburetor. While it's true that carburetors can be made from many different materials, the appropriate materials are severely restricted once mechanistic context and desired function are considered. A plastic carburetor from a lawn mower engine will cease to function as a carburetor in the context of a Ford F150 engine. The function and the context in which the function is embedded determine the range of structures that can implement that function. In the context of cognitive science, consider the function of storing information long term in a read/write, addressable form similar to the way memory works in digital computers (Gallistel and King 2009). Fulfilling this function requires memory registers whose states persist over a sufficiently long time, which must be appropriately connected to the processing components; it also requires a system of addresses that are stored in memory components and manipulated by an appropriate control structure. None of this comes for free by positing a certain function; for a functional hypothesis to prove correct, the structures that perform that function within the nervous system must be identified.

The upshot is that cognition cannot be explained without accounting for the ways in which structures constrain functions and vice versa. In the long run, the mutual constraints between structures and functions lead cognitive psychologists and neuroscientists to look to each other's work to inform their analyses. At any given level of organization, the goal is to identify both what structures are in play and what functions





Journal: 11229-SYNT Article No.: 0783 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

are performed. The more we know about functions and the context in which they are embedded, the more we can formulate sensible hypotheses about which structures must be involved. Similarly, the more we know about structures and the contextual factors that influence them, the more we can formulate sensible hypotheses about which functions they perform. The best strategy is to investigate *both* structures and functions simultaneously. As we will illustrate in the next section, this is the main driving force between the merging of neuroscience and cognitive psychology into cognitive neuroscience.

Building on these observations about the relations between structures and functions, similar points can be made about *implementation* (or realization): there is no single implementation (or realization) level. Every system of organized components implements (realizes) the capacities of the whole it composes. Every capacity of a whole is implemented (realized) by its organized components. Implementation is thus relative to level. In other words, every level of a multilevel mechanism is implementational relative to the level above it. The only exceptions to this occur at the (somewhat arbitrary) boundaries of cognitive neuroscientific inquiry—e.g., the whole behaving organism need not implement anything (at least as far as cognitive neuroscience is concerned).[11]

Relatedly, every level of a neurocognitive mechanism is *neural*—or more precisely, every level is either (at least partially) composed of neurons or is a component of a neuron. The fact that neurocognitive mechanisms “bottom out” in components of neurons is a contingent feature of the disciplinary boundaries of cognitive neuroscience. The crucial point for present purposes, in terms of deviation from the classical autonomist view, is that there is no “non-mechanistic” level of explanation to be added to the mechanistic ones.

Whether a level of a neurocognitive mechanism is *representational* or *computational* depends on whether it contains representations or performs computations (in accord with the above definitions of these terms). Many cortical areas and other large neural systems contain representations and perform computations in the relevant sense, so they are representational and computational. Many of their components (columns, nuclei) also contain representations and perform more limited computations over them; the computations they perform are component processes of the computations performed by their containing systems. Therefore, large neural components are representational and computational, and the same holds for their components (e.g. networks and circuits). Again, the computations performed by smaller components are constituents of the larger computations performed by their containing systems, and that

[11] Here we depart from Craver (2007, pp. 212ff.), who distinguishes between levels of mechanistic organization and levels of realization. Craver adopts the view that realization is a relation between two properties of one and the same whole system, not to be confused with the relation that holds between levels of mechanistic organization. (According to Craver, as according to us, levels of mechanistic organization are systems of components, their capacities, and their organizational relations, and they are related compositionally to other levels of mechanistic organization.) We reject the account of realization adopted by Craver; we hold that each level of mechanistic organization realizes the mechanistic level above it and is realized by the mechanistic level below it (Piccinini and Maley 2014). Realization, in its most useful sense, is precisely the relation that obtains between two adjacent mechanistic levels in a multi-level mechanism and is thus a compositional relation.



Journal: 11229-SYNT Article No.: 0783 TYPESET ☑ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

is how the computations of their containing systems are mechanistically explained. At a still lower level, the response profiles of some single neurons reliably correlate with specific variables and it appears to be their *function* to correlate in this way; if this is correct, then they are representational in the relevant sense. Whether individual neurons perform computations over these representations is a matter of debate that can be left open. Sub-neuronal structures may or may not contain representations and perform computations depending on the extent to which they satisfy the relevant criteria. At some point, we reach explanations that are no longer computational but instead are purely biophysical. Here certain biophysical mechanisms explain how certain neural systems register and transmit information.[12] These purely biophysical (and lower) levels are no longer representational and computational in the relevant sense.

Finally, whether a level of a neurocognitive mechanism is *cognitive* depends on whether and how it contributes to a cognitive capacity. Given our account in the previous section, to the effect that explaining cognitive capacities involves neural computation and representation, a neurocognitive level is cognitive depending on the extent to which the components of that level perform computations over representations in a way that is relevant to explaining some cognitive capacity. As above, the lowest-level neural computations are explained purely biophysically. In some simple organisms, these simple computations may be sufficient to explain the organism's behaviors. In more complex organisms, however, these simple computations combine with other simple neural computations to constitute higher level neural computations, which in turn constitute still higher level neural computations, and so on, until we reach the highest level neural computations, which explain cognitive capacities.

An example of such an explanatory strategy would be the following sketch of an account of vision. Individual cells in V1 selectively respond to particular line orientations from the visual scene. Several of these cells in conjunction form an orientation column, which provide the basis for edge detection in the visual scene. These orientation columns combine to constitute V1, which computes the boundaries of visual objects. V1 then operates in conjunction with downstream parietal and temporal areas to constitute the different "streams" of visual processing and visual object representation.[13]

The resulting framework for explaining cognition is a mechanistic version of homuncular functionalism, whereby higher-level cognitive capacities are iteratively explained by lower-level cognitive capacities until we reach a level at which the lower-level capacities are no longer cognitive in the relevant sense (Attneave 1961; Fodor 1968b; Dennett 1978; Lycan 1981; Maley and Piccinini 2013). The rise of cognitive neuroscience illustrates how this framework has developed and been applied (and continues to develop and be applied) in scientific practice. In the next section, we highlight three aspects of cognitive neuroscience that demonstrate the development and appli-

[12] The purely biophysical level is reached when our explanation of the processes no longer appeals solely to differences between different portions of the vehicles along relevant dimensions of variation—which in the case of neural vehicles are mostly spike frequency and timing—in favor of the specific biophysical properties of neurons, such as the flow of specific ions through their cell membranes.

[13] We are not committed to the adequacy of this particular explanation of visual processing, just to its exemplifying the explanatory strategy of iterated computational mechanisms that we are explicating here.



Journal: 11229-SYNT Article No.: 0783 TYPESET ✓ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

cation of this framework: the incorporation of experimental protocols from cognitive psychology into neuroscience experiments, the evolution of functional neuroimaging, and the movement toward biological realism in computational modeling in cognitive science.

## 6 How cognitive neuroscience exhibits multilevel mechanistic integration

Cognitive neuroscience emerged as a discipline in the late 1980s. Prior to that time, cognitive science and neuroscience had developed largely in isolation from one another. Cognitive science developed between the 1950s and the 1970s as an interdisciplinary field comprised primarily of aspects of psychology, linguistics, and computer science. In linguistics, this involved the development of generative grammars aimed at explaining the syntax structuring human linguistic behavior. In psychology, researchers began developing information processing accounts aimed at explaining capacities like problem solving and memory encoding. In computer science, researchers began developing computational models, involving discrete state-transitions, in order to model psychological capacities like reasoning and problem solving. The development of cognitive science accelerated through the 1960s and 1970s, with these approaches proceeding on their own terms with little contact with neuroscience. While during this period the hypothesis space for cognitive functions was constrained, the lack of contact with neuroscientific evidence contributed to a significant underdetermination of these hypotheses by available evidence (cf. Anderson 1978).

Meanwhile, neuroscience developed as an interdisciplinary field investigating both normal and abnormal functioning of the nervous system. Neurophysiological investigations had been carried out since at least the 1890s, at a time when neuroscience and psychology were seen as disciplines that should be integrated (e.g., Freud 1895/1966; James 1890/1983). But the term "neuroscience" was only coined in the 1960s with the development of new techniques for investigating the cellular and molecular levels of nervous systems and for relating those investigations to systems and behavioral levels. As a result, early neuroscience illuminated candidate structures for implementing cognitive functions, but it did so with little connection to functional context, thereby making limited progress towards explaining cognitive functions.

Throughout the development of both fields in the 1960s and 1970s, neuroscience and cognitive science dealt with domains with a great degree of overlap. In principle, they could have merged; in practice, they tended to exclude one another. Conceptual motivation for this exclusion was rooted in views already discussed: autonomist commitments (both implicit and explicit) among practicing cognitive scientists versus reductionist commitments among many practicing neuroscientists. Meanwhile, practical motivation that reinforced the exclusion was rooted in the pace of early developments that shaped both fields. In neuroscience, techniques for investigation at the cellular and molecular level developed at a pace that outstripped and overshadowed work at the systems level. In cognitive science, rapid developments in computer science and artificial intelligence in the 1970s provided a computational framework in which processes were decomposed into specific operations performed on symbolic (language-like) structures. This framework fostered a gulf between cognitive analyses



Journal: 11229-SYNT Article No.: 0783 TYPESET ☑ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

and neural analyses because there was no obvious way for these symbolic computational units to be realized in neural tissue.

The differences between the fields began to abate in the 1980s. Bechtel (2001) cites two chief contributors: the need for more sophisticated behavioral protocols in neuroscience, and the related development of functional neuroimaging techniques.

The former developed naturally as neuroscience researchers shifted focus toward determining specific functions performed by recently discovered cellular and molecular structures, attempting to link those structures to particular behaviors. In order to draw these links and target higher-level functions, neuroscientists needed more sophisticated behavioral protocols. Cognitive psychologists had developed relatively sophisticated behavioral protocols in order to obtain informative data from a limited range of dependent variables. At the time, most experiments in cognitive psychology involved inference to some cognitive hypothesis based on patterns in two dependent variables: characteristic patterns of error human subjects exhibited on some task (error rate), and the typical amount of time taken for those subjects to perform the task (reaction time).

As neuroscientists began to shift their explanatory ambitions, they ran up against the same limited range of dependent variables. Rather than reinventing the wheel, they began incorporating behavioral protocols from cognitive psychology and applying those protocols to experimental setups in which neural activity could be monitored in both humans and model organisms. These more sophisticated behavioral protocols allowed neuroscientists to form and test hypotheses about the contributions of cellular and molecular structures to higher-level functions.

This disciplinary shift demonstrates how, in practice, functions constrain structures: sophisticated behavioral protocols provided the functional context necessary to constrain the search for the structures involved in performing those functions. Of course, many of these protocols were subsequently revised in a give-and-take between the incoming physiological data and the existing functional models that motivated the protocols. But with the integration of these techniques and protocols, the underdetermination of structure-function mapping became a tractable empirical issue rather than a conceptual one.

The other main contributor to the practical integration of psychology and neuroscience has been the development of functional neuroimaging techniques, which allow measurement of physiological changes in large neural structures in response to performance of particular tasks. The first functional neuroimaging technique to be developed was Positron Emission Tomography (PET). PET involves injecting a radioactive tracer into a subject's bloodstream, which can then be imaged as it decays to illuminate blood flow to different brain regions. In a seminal study, Fox et al. (1986) used PET to measure hemodynamic response in particular brain areas during different cognitive tasks—their results correlated sensory and motor tasks with increased blood flow in primary sensory and motor areas, respectively.

Prior to the development of neuroimaging, the primary way to attribute specific cognitive functions to neural systems and thereby to relate neural activity to behavior (in humans) was through the study of behavioral deficits resulting from lesions due to some form of traumatic brain injury. While these lesion studies remain an integral part of cognitive neuroscience to this day, there are a number of potential confounding



Journal: 11229-SYNT Article No.: 0783 TYPESET ☑ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

factors involved in extrapolating from these data to brain function in non-pathological cases (see, e.g., Kosslyn and Van Kleeck 1990). Brain imaging assuages some of these concerns, and as a result the early research into the applications of PET for functional brain imaging set the stage for the explosion of research in cognitive neuroscience precipitated by the development of even more powerful and noninvasive imaging techniques like functional Magnetic Resonance Imaging (fMRI).

The ability to correlate activity in different brain regions with specific tasks has improved our ability to map cognitive functions to neural structures. Underdetermination problems remain, as cognitive functions cannot be simply read off of tasks and functional neuroimaging still has limits on spatial and temporal resolution that place corresponding limits on fine-grained attribution of functions to lower-level neural structures (see, e.g., Roskies 2009 for further discussion). Nonetheless, these neuroimaging techniques provide valuable data to constrain structure-function mapping by situating putative functions within mechanistic context.

This mechanistic context needs to be supplemented by further modeling in order to provide fully integrated explanations of how cognitive phenomena relate to neural activity. For instance, the recent trend toward model-based fMRI studies, in which models from computational neuroscience are incorporated into traditional fMRI experimental designs, demonstrates one way in which these integrated explanations are currently being approached (e.g., O'Doherty et al. 2007; cf. Egan and Matthews 2006; Povich forthcoming). These model-based imaging techniques illustrate, with particular clarity, the applications of the multilevel mechanistic framework we have been advancing. At a relatively coarse-grained level, neuroimaging allows identification of the cortical and subcortical networks that are active in particular tasks. In order to determine more precisely the functions performed by these intermediate-level networks, researchers look to modeling efforts in computational neuroscience that are highly constrained by the neurophysiology of the particular regions involved (more on this below). This strategy facilitates integration between different mechanistic levels and in so doing allows more precise identification of the functions involved in cognitive processes and the specific structures performing those functions. The proliferation of neuroimaging studies over the past two decades and, in particular, the current trend toward model-based approaches provide further evidence that cognitive neuroscience is indeed a science concerned with the complexities of structure-function mapping, rather than a science predicated on giving primacy to one over the other.

Finally, the evolution of computational modeling in cognitive science also exemplifies the shift from autonomist cognitive science to cognitive neuroscience. After McCulloch and Pitts (1943) introduced the first model of neural computation, three main modeling research programs developed. First, there is classical symbolic computationalism, which strives to explain cognitive capacities in terms of symbolic computation in putative autonomy from neuroscience (e.g., Newell and Simon 1972; Anderson 1983). Second is connectionism, which strives to explain cognitive capacities in terms of neural network computations, though such neural networks are artificial models that are minimally (if at all) constrained by what is known about actual neural systems (e.g., Rosenblatt 1962; Feldman and Ballard 1982). The third modeling research program is computational neuroscience, which explains cognitive capacities by building models of neural systems that are explicitly constrained by known





Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

neuroanatomical and neurophysiological evidence (e.g., Hodgkin and Huxley 1952; Caianiello 1961; Stein 1965; Knight 1972; Wilson and Cowan 1972). The critical difference is that while classicist and connectionist models cannot be mapped onto neural structures in any direct way, models from computational neuroscience target specific neural structures and form hypotheses about the specific functions they perform and how those functions contribute to cognitive behaviors (cf. Kaplan 2011). Thus, computational neuroscience models exhibit the integration of functions and structures that we have argued characterizes cognitive neuroscience.

For much of their history, these three traditions developed largely independently from one another. Classical computationalism gained a solid footing in the 1970s and was based on the idea (outlined above in Sect. 2) that the brain is a universal, program-controlled digital computer. The idea behind this modeling paradigm was that cognitive processes can be seen as the software that is implemented on such computers and thus can be studied independently from the hardware/wetware implementing the software. But this analogy between natural cognitive systems and digital computers is problematic for two reasons. First, whether nervous systems are universal, program-controlled digital computers is an empirical question; such a question cannot be settled independently of neuroscience. Further, and more importantly, evidence from neuroscience suggests that neural computation, in the general case, is in fact *not* a form of digital computation (Piccinini and Bahar 2013). Since digital computation is a necessary (though insufficient) condition for a system to be a universal, program-controlled digital computer, current best evidence suggests that nervous systems are in fact not such systems.

In the 1980s, connectionism re-emerged and challenged the hardware/software analogy in favor of "neurally inspired" network models (Rumelhart et al. 1986, p. 131). But typical neo-connectionist psychology was not grounded in known neural processes and mechanisms. Connectionists made largely arbitrary assumptions about the number of neurons, number of layers, connectivity between neurons, response properties of neurons, and learning methods. Connectionist psychology made such assumptions in order to model and explain psychological phenomena. Since these assumptions were not grounded in neuroscience, connectionists were merely developing a different take on the standard computer analogy, replete with their own commitment to the autonomy of psychology from neuroscience. Thus, while connectionism pushed in the right direction, it fell short of actually integrating psychology and neuroscience. From the point of view of cognitive neuroscience, this kind of connectionism was more on the side of classical, autonomist cognitive science than it was on the side of neuroscience. As a result, both classical computationalism and connectionism foster models of cognitive systems that are autonomous from structural (neuroscientific) constraints (cf. Weiskopf 2011).

While philosophers were captivated by the divide between classical computationalism and connectionism, computational neuroscientists developed powerful tools for modeling and explaining cognitive phenomena in terms of actual biological processes. They imported theoretical tools from mathematics and physics and took advantage of the exponentially increasing power of modern computers. By now, there are many highly sophisticated research programs developing detailed models of how specific neural structures perform cognitive functions at various levels of organization (e.g., Dayan and Abbott 2001; Ermentrout and Terman 2010; Eliasmith 2013).



Journal: 11229-SYNT Article No.: 0783 TYPESET ☑ DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

The field has matured to a point where connectionism is disappearing as an independent research tradition, instead merging into computational cognitive neuroscience (O'Reilly and Munakata 2000; O'Reilly et al. 2014). Most of the classicist research programs are also being shaped by this emergence of computational neuroscience. While we lack space for more detailed treatment, recent pronouncements of some key figures in classical and connectionist modeling indicate that the field is undergoing a deep transformation.

Early attempts at building classical cognitive architectures were based on *production systems* (Anderson 1983; Laird and Newell 1987). Production systems model cognitive processes as software packages specifying a series of "if… then…" statements (rules) taking inputs to outputs. Initially, these quintessentially symbolic models were in no way constrained by neural data. Nevertheless, their proponents expressed great confidence: "Cognitive skills are realized by production rules. This is one of the most astounding and important discoveries in psychology and may provide a base around which to come to a general understanding of human cognition" (Anderson 1993, p. 1). More recently, work on these same cognitive architectures has evolved to respect multiple levels of computational organization that are constrained by evidence from neuroscience. A stark transition can be seen, in particular, in Anderson's work, where his initial ambitions for his ACT-R production system architecture (as a univocal model for cognition) have been replaced by the acknowledgement that hybrid architectures are more promising. In a recent paper, Anderson et al. argue that "theories at different levels of detail and from different perspectives are mutually informative and constraining, and furthermore no single level can capture the full richness of cognition" (Jilk et al. 2008). Similarly, Laird's most recent presentation of his Soar architecture advocates constraint by evidence from neuroscience (as well as from psychology and AI): "we have found that combining what is known in psychology, in neuroscience, and in AI is an effective strategy to building a comprehensive cognitive architecture" (Laird 2012, p. 22).

Similar transitions can be seen in the works of other leading cognitive scientists. Stephen Kosslyn, a pioneer of mental imagery and the view that mental imagery involves a special, pictorial representational format, went from a traditional theory based primarily on behavioral data (Kosslyn 1980) to a thoroughly cognitive neuroscientific framework (Kosslyn 1994; Kosslyn et al. 2006). Kosslyn's trajectory is a good illustration of the process of deepening explanations via the investigation of underlying mechanisms (Thagard 2007), which is the hallmark of cognitive neuroscience. Kosslyn's early theory of mental imagery faced skeptical resistance from defenders of a non-pictorial alternative (Pylyshyn 1981). By appealing to fMRI and neuropsychological evidence, Kosslyn later gained widespread acceptance for his pictorial theory. The debate over the format of mental images is not entirely over, but the way to resolve it is not to reject neuroscientific evidence as irrelevant or insufficient (cf. Pylyshyn 2002, 2003). The way to resolve it is to learn even more about how the brain realizes and processes mental images.

An analogous shift from traditional cognitive science to cognitive *neuro*science can be seen in Anne Treisman's landmark work on attention (Treisman and Gelade 1980; Treisman 1996, 2009). James McClelland, who pioneered the neo-connectionist models that were developed autonomously from neuroscience (Rumelhart et al. 1986),



Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

subsequently co-founded the Center for the Neural Basis of Cognition (a collaboration between Carnegie-Mellon University and the University of Pittsburgh) and has become a computational cognitive neuroscientist (e.g., McClelland and Lambon Ralph 2013). Michael Posner's authoritative treatment of the subtractive method employed in cognitive psychology (Posner 1976) became the basis for the rigorous use of neuroimaging methods, beginning with PET, that are the backbone of much cognitive neuroscience (Posner and Raichle 1994).

Because we are still in the midst of this interdisciplinary shift toward the integration of psychology and neuroscience, it is easy to miss how revolutionary it is. The old view of psychology as autonomous from neuroscience (as well as the faith in the reductionist program, from the other direction) has been effectively supplanted by a new framework where multilevel integration rules the day.

## 7 Conclusion

The cognitive neuroscience revolution consists in rejecting the scientific practices stemming from the traditional two-level view of cognitive science and replacing them with a fully integrated science of cognition. The traditional two-level view maintained a division of labor between the sciences of cognition proper (psychology, linguistics, anthropology, AI, and philosophy) and sciences of implementation (neuroscience). This framework has fallen by the wayside as cognitive neuroscience has risen to prominence.

The old two-level picture fell apart for several reasons. First, new modeling and empirical techniques—including the emergence of neuroimaging methods—have provided more sophisticated ways to link cognitive capacities to the activities of specific neural systems. Second, the dubious assumptions about the nervous system that bolstered the received view, such as the assumption that the nervous system is a universal, program-controlled digital computer, simply have not panned out. Third, the received view of cognitive explanation, according to which there is one privileged cognitive level and one distinctive and autonomous explanatory style—functional analysis—has turned out to be faulty.

We have argued that philosophy of cognitive science should take heed. In place of the eliminative/reductive and classical functionalist/autonomist views of cognitive science, we have proposed the framework of integrated, multilevel, representational, and computational neural mechanisms as capturing the essence of successful explanation in cognitive neuroscience. Any discipline that studies cognition can fruitfully contribute to this project by characterizing one or more neurocognitive level(s) using the various empirical and analytical techniques at its disposal. In addition to avoiding the problems of the old two-level view, this framework also avoids the pitfalls of both reduction and elimination by retaining a role for organization within each neurocognitive level. While much work remains to be done in order to more fully understand the implications, applications, and limitations of this framework, the first step lies in accepting the revolutionary shift in our understanding of the physical bases of cognition that has already taken place.

## References


Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*, *85*, 249–277.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Attneave, F. (1961). In defense of homunculi. In W. Rosenblith (Ed.), *Sensory communication* (pp. 777–782). Cambridge, MA: MIT Press.

Barberis, S. D. (2013). Functional analyses, mechanistic explanations, and explanatory tradeoffs. *Journal of Cognitive Science*, *14*(3), 229–251.

Barrett, D. (2014). Functional analysis and mechanistic explanation. *Synthese*. doi:10.1007/s11229-014-0410-9.

Batterman, R. (2013). The tyranny of scales. In R. W. Batterman (Ed.), *The Oxford handbook of philosophy of physics* (pp. 255–286). New York: oxford University Press.

Bechtel, W. (2015). Investigating neural representations: The tale of place cells. *Synthese*.

Bechtel, W. (2001). Cognitive neuroscience: Relating neural mechanisms and cognition. In P. Machamer, P. McLaughlin, & R. Grush (Eds.), *Philosophical reflections on the methods of neuroscience*. Pittsburgh, PA: University of Pittsburgh Press.

Bechtel, W. (2008). *Mental mechanisms: Philosophical Perspectives on cognitive neuroscience*. London: Routledge.

Bechtel, W., & Abrahamsen, A. (2013). Thinking dynamically about biological mechanisms: Networks of coupled oscillators. *Foundations of Science*, *18*, 707–723.

Bechtel, W., & Shagrir, O. (forthcoming). The non-redundant contributions of Marr's three levels of analysis for explaining information processing mechanisms. *Topics in Cognitive Science*.

Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer.

Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience. *Synthese*, *151*, 411–434.

Bogen, J. (2005). Regularities and causality: Generalizations and causal explanations. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 397–420.

Boone, W., & Piccinini, G. (unpublished). Mechanistic abstraction.

Burge, T. (2010). *Origins of objectivity*. Oxford: Oxford University Press.

Caianiello, E. R. (1961). Outline of a theory of thought processes and thinking machines. *Journal of Theoretical Biology*, *1*(2), 204–235.

Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of Science*, *75*, 1–27.

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese*, *191*(2), 127–154.

Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*, *78*, 67–90.

Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain*. Cambridge, MA: MIT Press.

Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

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. In K. F. C. & W. R. A. (Eds.), *Explanation and cognition*. Cambridge: Cambridge University Press.

Daugman, J. G. (1990). Brain metaphor and brain theory. In E. L. Schwartz (Ed.), *Computational neuroscience* (pp. 9–18). Cambridge, MA: MIT Press.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.

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

Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Dretske, F. I. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Egan, F., & Matthews, R. (2006). Doing cognitive neuroscience: A third way. *Synthese*, *153*, 377–391.

Ermentrout, G. B., & Terman, D. H. (2010). *Mathematical foundations of neuroscience*. New York: Springer.






Journal: 11229-SYNT Article No.: 0783 TYPESET DISK LE CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford: Oxford University Press.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, *6*, 205–254.

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. In J. Tomberlin (Ed.), *Philosophical perspectives 11: Mind, causation, and world* (pp. 149–163). Boston: Blackwell.

Fodor, J. A. (1998). *Concepts*. Oxford: Clarendon Press.

Fox, P. T., Minton, M. A., Raichle, M. E., Miezin, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping human visual cortex with positron emission tomography. *Nature*, *323*, 806–809.

Fresco, N. (2014). *Physical computation and cognitive science*. New York: Springer.

Freud, S. (1895/1966). Project for a scientific psychology. In E. Jones (Ed.) & J. Strachey (Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 1, pp. 295–397). London: Hogarth Press.

Gallistel, R. G., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will transform neuroscience*. New York: Wiley/Blackwell.

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*(3), S342–S353.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. New York: Oxford University Press.

James, W. (1890/1983). The principles of psychology. Cambridge, MA: Harvard University Press.

Jilk, D., Lebiere, C., O'Reilly, R., & Anderson, J. (2008). SAL: An explicitly pluralistic cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, *20*(3), 197–218.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. New York: Cambridge University Press.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese, 183*(3), 339–373.

Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical models. *Philosophy of Science*, *78*(4), 601–627.

Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *Journal of General Physiology*, *59*, 734–766.

Kosslyn, S. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Kosslyn, S., & Van Kleeck, M. H. (1990). Broken brains and normal minds: Why humpty–dumpty needs a skeleton. In E. L. Schwartz (Ed.), *Computational neuroscience*. Cambridge, MA: MIT Press.

Kosslyn, S., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford University Press.

Laird, J. E. (2012). *The soar cognitive architecture*. Cambridge, MA: MIT Press.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence*, *33*, 1–64.

Levy, A. (2013). What was Hodgkin and Huxley's achievement? *British Journal for Philosophy of Science*. doi:10.1093/bjps/axs043.

Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science*, *80*(2), 241–261.

Lycan, W. (1981). Form, function, and feel. *Journal of Philosophy*, *78*, 24–50.

Lycan, W. (1990). The continuity of levels of nature. In W. Lycan (Ed.), *Mind and cognition*. Malden, MA: Blackwell.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25.

Maley, C., & Piccinini, G. (2013). Get the latest upgrade: Functionalism 6.3.1. *Philosophia Scientiae*, *17*(2), 135–149.

Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman and Company.

McClelland, J. L., & Lambon Ralph, M. A. (Eds.). (2013). *Cognitive neuroscience: Emergence of mind from brain. The biomedical & life sciences collection*. London: Henry Stewart Talks Ltd.

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

Milkowski, M. (2013). *Explaining the computational mind*. Cambridge, MA: MIT Press.

Morgan, A. (2014). Representations gone mental. *Synthese*, *191*(2), 213–244.

Newell, A. (1980). Physical symbol systems. *Cognitive Science*, *4*, 135–183.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.

Newell, A., & Simon, H. A. (1976). Computer science as an empirical enquiry: Symbols and search. *Communications of the ACM*, *19*, 113–126.

O'Doherty, J., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning and decision making. *Annals of the New York Academy of Sciences*, *1104*, 35–53.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain*. Cambridge, MA: MIT Press.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2014). *Computational cognitive neuroscience*. Wiki Book (2nd ed.). http://ccnbook.colorado.edu.

Piccinini, G. (unpublished). Activities are manifestations of causal powers.

Piccinini, G. (2007). Computational modeling vs. computational explanation: Is everything a turing machine, and does it matter to the philosophy of mind? *Australasian Journal of Philosophy*, *85*(1), 93–115.

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

Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive Science*, *34*, 453–488.

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311.

Piccinini, G., & Maley, C. (2014). The metaphysics of mind and the multiple sources of multiple realizability. In M. Sprevak & J. Kallestrup (Eds.), *New waves in the philosophy of mind*. New York: Palgrave Macmillan.

Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of Biological Physics*, *37*(1), 1–38.

Posner, M. I. (1976). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: Scientific American Books.

Povich, M. (forthcoming). Mechanisms and model-based fMRI. *Philosophy of Science*.

Putnam, H. (1967). *Psychological predicates. Art, philosophy, and religion*. Pittsburgh, PA: University of Pittsburgh Press.

Putnam, H. (1975). Philosophy and our mental life. In H. Putnam (Ed.), *Mind, language and reality: Philosophical papers* (Vol. 2, pp. 291–303). Cambridge: Cambridge University Press.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*, *88*, 16–45.

Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, *25*(2), 157–237.

Pylyshyn, Z. W. (2003). Return of the mental image: Are there really pictures in the head? *Trends in Cognitive Science*, *7*(3), 113–118.

Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Washington, DC: Spartan.

Roskies, A. (2009). Brain–mind and structure–function relationships: A methodological response to Coltheart. *Philosophy of Science*, *76*(5), 927–939.

Rumelhart, D. E., McClelland, J. M., & The PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Rusanen, A.-M., & Lappi, O. (2007). The limits of mechanistic explanation in cognitive science. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the European cognitive science conference 2007* (pp. 284–289). Howe: Lawrence Erlbaum Associates.

Scarantino, A. (2015). Information as a probabilistic difference maker. *Australian Journal of Philosophy*. doi:10.1080/00048402.2014.993665.

Schaffner, K. F. (2008). Theories, models, and equations in biology: The heuristic search for emergent simplifications in neurobiology. *Philosophy of Science*, *75*, 1008–1021.





Journal: 11229-SYNT Article No.: 0783 ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:2015/6/5 Pages: 26 Layout: Small-X

Shagrir, O. (2010a). Brains as analog-model computers. *Studies in History and Philosophy of Science*, *41*, 271–279.

Shagrir, O. (2010b). Marr on computational-level theories. *Philosophy of Science*, *77*, 477–500.

Smart, J. J. C. (1959). Sensations and brain processes. *The Philosophical Review*, *68*(2), 141–156.

Smart, J. J. C. (2007). The mind/brain identity theory. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2007 ed.). http://plato.stanford.edu/archives/sum2007/entries/mind-identity/.

Stein, R. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal*, *5*(2), 173–194.

Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*, *74*, 28–47.

Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, *6*(2), 171–178.

Treisman, A. (2009). Attention: Theoretical and psychological perspectives. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 189–204). Cambridge, MA: MIT Press.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Van Eck, D., & Weber, E. (2014). Function ascription and explanation: Elaborating an explanatory utility desideratum for ascriptions of technical functions. *Erkenntnis*. doi:10.1007/s10670-014-9605-1.

Vartanian, A. (1973). In P. P. Wiener (Ed.), *Dictionary of the history of ideas: Studies of selected pivotal ideas*. New York: Scriners.

Waxman, S. (1972). Regional differentiation of the axon: A review with special reference to the concept of the multiplex neuron. *Brain Research*, *47*, 269–288.

Weber, M. (2005). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.

Weber, M. (2008). Causes without mechanisms: Experimental regularities, physical laws, and neuroscientific explanation. *Philosophy of Science*, *75*(5), 995–1007.

Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford University Press.

Weiskopf, D. (2011). Models and mechanisms in psychological explanation. *Synthese*, *183*(3), 313–338.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, *12*, 1–24.

Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: University of Chicago Press.






Journal: **11229-SYNT** Article No.: **0783** ☐ TYPESET ☑ DISK ☐ LE ☐ CP Disp.:**2015/6/5** Pages: **26** Layout: **Small-X**