

The cognitive neuroscience revolution

Worth Boone¹ · Gualtiero Piccinini²

Received: 22 September 2014 / Accepted: 22 May 2015
© Springer Science+Business Media Dordrecht 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

16 **1 Introduction**

17 The traditional framework of cognitive science included (aspects of) six disciplines:
18 psychology, computer science, linguistics, anthropology, neuroscience, and philos-
19 ophy. These six disciplines were supposed to work together towards understanding
20 cognition in accordance with a neat division of labor, to which many practitioners
21 conformed. On one side stood psychology, with the help of computer science, linguis-
22 tics, anthropology, and philosophy; on the other side stood neuroscience. Psychology
23 etc. studied the functional or cognitive level, or—in Marr’s terminology—the compu-
24 tational and algorithmic levels; neuroscience investigated the neural, mechanistic, or
25 implementation level. Explanations at these two levels were considered distinct and
26 autonomous from one another.

27 This division of labor leaves no room for cognitive *neuroscience*. Indeed, from
28 this perspective, the very term “cognitive neuroscience” is almost an oxymoron,
29 because neuroscience is supposed to deal with the mechanisms that implement cog-
30 nitive processes, not with cognition proper. Yet cognitive neuroscience has emerged
31 as the new mainstream approach to studying cognition. What gives?

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

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

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

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

77 2 Cognitive science as traditionally conceived

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

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

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

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

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

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

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

138 [I]n the language of neurology ..., presumably, notions like computational state
 139 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.

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

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

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

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

152 3 Traditional responses to cognitive science

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

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

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

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

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

184 Contemporary cognitive neuroscience aims to explain cognition on the basis of
185 neural computation over neural representations (more on this below). If the elimi-
186 nativist approach implies that cognition itself—and all “cognitive” theoretical posits,
187 such as representation, computation, or information processing—should be eliminated
188 or at least deflated (cf. Ramsey 2007), then we are faced with a solution that is anti-
189 thetical to cognitive neuroscience.

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

199 Nevertheless, some have argued that when we can intervene on molecular struc-
200 tures in the brain and affect some cognitive behavior, the specific molecular events
201 “directly explain” the behavioral data, we thereby reduce the relevant cognitive capac-
202 ity to the relevant molecular events, and we thereby obviate the need for intermediate
203 levels of explanation (cf. Bickle 2003, 2006). The main problem with this form of
204 reductionism is that specific molecular events are at best only *partial* explanations
205 of cognitive phenomena. It is one thing to correlate specific molecular events with
206 cognitive phenomena via some specific intervention; to actually explain a cognitive
207 phenomenon on the basis of molecular events requires determining the ways in which
208 the molecular events are causally relevant to the production of the phenomenon of
209 interest. Molecular events are only relevant to the extent that they occur within spec-
210 ific neural structures, and locating the relevant neural structures requires more than
211 purely molecular neuroscience. In addition, even identifying a molecular event within
212 a neural structure that contributes to a cognitive behavior falls short of a full expla-
213 nation. A full explanation requires identifying how molecular events contribute to
214 relevant neural events, how relevant neural events contribute to circuit and network
215 events, how those in turn contribute to relevant systems-level events, and finally how
216 the relevant systems, appropriately coupled with the organism's body and environ-
217 ment, produce the behavior. These intermediate links in the causal-mechanistic chain
218 are crucial to connecting molecular events to cognitive phenomena in a way that is
219 explanatory, as opposed to merely correlational. And identifying these intermediate
220 level structures and their causal contributions requires going well beyond molecular
221 neuroscience (cf. Craver 2007; Bechtel 2008).

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

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

230 4 Multilevel neurocognitive mechanisms

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

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

258 Multilevel neurocognitive mechanisms have an iterative structure: at any level,
 259 each component of the mechanism is in turn another mechanism whose capacities
 260 are explained by the organized capacities of *its* components; and each whole mech-
 261 anism is itself a component part that contributes to the capacities of a larger whole.
 262 This multilevel iterative structure tops off in the capacities of whole organisms and
 263 their interactions with other organisms, which are studied by social neuroscience
 264 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).

265 neurotransmitters—that fall outside the disciplinary boundaries of cognitive neuro-
266 science.

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

273 A vehicle carries *semantic information* about a source just in case it reliably cor-
274 relates with the states of the source ([Dretske 1981](#); [Piccinini and Scarantino 2011](#);
275 [Scarantino 2015](#)). For instance, the spike trains generated by neurons in cortical area
276 V1 reliably correlate with the presence and location of edges in the visual environ-
277 ment; thus, they carry semantic information about the presence and location of edges
278 in the visual environment. But correlation alone is insufficient for representation.

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

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

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

305 The above account of computation is diametrically opposed to persistent views of
306 computation that draw a stark contrast between computational and mechanistic expla-
307 nations.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

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.

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

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

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

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

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

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

407 Special cases of this type of mechanism sketch are descriptions of computational
 408 (the function computed and why it is adequate to the task, cf. [Shagrir 2010b](#)) and
 409 algorithmic (the computational operations and representations) levels of a system,
 410 which omit details about the components that carry out the algorithm. There is cer-
 411 tainly value to such approaches in cognitive neuroscience, particularly in the context
 412 of discovery. [Marr \(1982\)](#) argued that a neural “algorithm” is “likely to be understood
 413 more readily by understanding the nature of the problem being solved than by exam-
 414 ining the mechanism (and the hardware) in which it is embodied” ([Marr 1982](#), p. 27).
 415 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.

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

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

436 5 Neurocognitive levels

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

444 Contrary to the received view, there is no single “functional,” “cognitive,” or “repre-
 445 sentational/computational” level of explanation, standing in opposition to a single (or
 446 even multiple) “structural,” “neural,” or “implementational” level(s).¹⁰ In this section
 447 we analyze each of these concepts from the perspective of neurocognitive mecha-
 448 nisms in order to highlight how our integrationist framework improves upon traditional
 449 autonomist and reductionist views.

450 In the first place, every level of a multilevel mechanism is both *functional* and
 451 *structural*, because every level contains structures performing functions. This stands in
 452 stark contrast to traditional views that maintain that structural analyses and functional
 453 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.

454 e.g. type physicalists (Smart 1959)—strove to identify mental types with physical
455 types. As a result, they may be interpreted as focusing on structural properties at
456 the expense of functional properties, relegating the latter to “second order states”
457 of physical types (Smart 2007). Traditional functionalists do the opposite: they give
458 primacy to functional properties at the expense of structural properties (e.g., Putnam
459 1967; Fodor 1968a). This is a somewhat unorthodox way of characterizing these views,
460 so some brief unpacking is in order.

461 Classical reductionist views of the mind-brain relation, specifically type identity
462 theorists, look to identify higher-order kinds (e.g. mental kinds, like “pain”) with
463 corresponding physical kinds. These reductive views were developed in contrast to
464 dualism: the view that the mind and brain are distinct kinds of substance. Dualists have
465 a notoriously difficult time specifying the means by which these distinct substances
466 interact; type identity theorists provide a dissolution of this problem. To say that water
467 is H₂O just is to identify a higher-level kind with a lower-level physical kind—an
468 arrangement of atoms. With such an identity in hand, it is illegitimate to wonder how
469 water and H₂O interact. In a similar vein, type identity theorists argued that mental
470 kinds, like pain, could be identified with particular neural kinds, like C-fibers firing.
471 This identification dissolves dualistic concerns about how mental states interact with
472 bodily states. What is noteworthy for present purposes is that the defining features of the
473 kinds that figure into higher-level analyses just are the lower-level physical features
474 common to instances of those kinds. This identification with lower-level physical
475 features downplays the role of functional features of those higher-level kinds. Water
476 is not individuated by its ability to quench thirst, nourish plants, etc., nor pain with its
477 role in avoidance of noxious stimuli, protecting injured body parts, etc. Instead both
478 kinds are identified with particular physical types, which possess the relevant causal
479 powers that are, incidentally, associated with these functions.

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

499 Neither of these approaches adequately captures the main thrust of work in cogni-
500 tive neuroscience, because that work is aimed at understanding the complex interplay
501 *between* structure and function. By contrast, the multilevel mechanistic framework we
502 are advocating adequately captures this aspect of cognitive neuroscientific explana-
503 tions; in our framework, functions constrain structures and vice versa. Functions *cum*
504 contextual factors—i.e. mechanistic context—constrain the range of structures that
505 can perform those functions. Similarly, structures *cum* contextual factors determine
506 the range of functions those structures can perform. In this framework, neither struc-
507 tures nor functions are given primacy over the other; neither can explain cognition
508 without the other.

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

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

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

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

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

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

570 Whether a level of a neurocognitive mechanism is *representational* or *compu-*
 571 *tational* depends on whether it contains representations or performs computations
 572 (in accord with the above definitions of these terms). Many cortical areas and other
 573 large neural systems contain representations and perform computations in the rele-
 574 vant sense, so they are representational and computational. Many of their components
 575 (columns, nuclei) also contain representations and perform more limited computations
 576 over them; the computations they perform are component processes of the computa-
 577 tions performed by their containing systems. Therefore, large neural components are
 578 representational and computational, and the same holds for their components (e.g. net-
 579 works and circuits). Again, the computations performed by smaller components are
 580 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.

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

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

604 An example of such an explanatory strategy would be the following sketch of an
605 account of vision. Individual cells in V1 selectively respond to particular line orienta-
606 tions from the visual scene. Several of these cells in conjunction form an orientation
607 column, which provide the basis for edge detection in the visual scene. These orienta-
608 tion columns combine to constitute V1, which computes the boundaries of visual
609 objects. V1 then operates in conjunction with downstream parietal and temporal areas
610 to constitute the different "streams" of visual processing and visual object represen-
611 tation.¹³

612 The resulting framework for explaining cognition is a mechanistic version of
613 homuncular functionalism, whereby higher-level cognitive capacities are iteratively
614 explained by lower-level cognitive capacities until we reach a level at which the lower-
615 level capacities are no longer cognitive in the relevant sense (Attneave 1961; Fodor
616 1968b; Dennett 1978; Lycan 1981; Maley and Piccinini 2013). The rise of cognitive
617 neuroscience illustrates how this framework has developed and been applied (and con-
618 tinues to develop and be applied) in scientific practice. In the next section, we highlight
619 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.

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

624 **6 How cognitive neuroscience exhibits multilevel mechanistic integration**

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

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

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

663 and neural analyses because there was no obvious way for these symbolic computa-
664 tional units to be realized in neural tissue.

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

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

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

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

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

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

factors involved in extrapolating from these data to brain function in non-pathological cases (see, e.g., [Kosslyn and Van Kleeck 1990](#)). Brain imaging allays 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

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

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

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

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

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

808 Early attempts at building classical cognitive architectures were based on *produc-*
809 *tion systems* (Anderson 1983; Laird and Newell 1987). Production systems model
810 cognitive processes as software packages specifying a series of “if... then...” state-
811 ments (rules) taking inputs to outputs. Initially, these quintessentially symbolic models
812 were in no way constrained by neural data. Nevertheless, their proponents expressed
813 great confidence: “Cognitive skills are realized by production rules. This is one of
814 the most astounding and important discoveries in psychology and may provide a base
815 around which to come to a general understanding of human cognition” (Anderson
816 1993, p. 1). More recently, work on these same cognitive architectures has evolved to
817 respect multiple levels of computational organization that are constrained by evidence
818 from neuroscience. A stark transition can be seen, in particular, in Anderson’s work,
819 where his initial ambitions for his ACT-R production system architecture (as a univocal
820 model for cognition) have been replaced by the acknowledgement that hybrid archi-
821 tectures are more promising. In a recent paper, Anderson et al. argue that “theories at
822 different levels of detail and from different perspectives are mutually informative and
823 constraining, and furthermore no single level can capture the full richness of cognition”
824 (Jilk et al. 2008). Similarly, Laird’s most recent presentation of his Soar architecture
825 advocates constraint by evidence from neuroscience (as well as from psychology and
826 AI): “we have found that combining what is known in psychology, in neuroscience,
827 and in AI is an effective strategy to building a comprehensive cognitive architecture”
828 (Laird 2012, p. 22).

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

843 An analogous shift from traditional cognitive science to cognitive neuroscience
844 can be seen in Anne Treisman’s landmark work on attention (Treisman and Gelade
845 1980; Treisman 1996, 2009). James McClelland, who pioneered the neo-connectionist
846 models that were developed autonomously from neuroscience (Rumelhart et al. 1986),

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

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

859 7 Conclusion

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

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

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

889 **References**

- 890 Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review*,
891 85, 249–277.
- 892 Anderson, J. R. (1983). *The architecture of cognition*. Cambridge: Harvard University Press.
- 893 Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.
- 894 Attneave, F. (1961). In defense of homunculi. In W. Rosenblith (Ed.), *Sensory communication* (pp. 777–
895 782). Cambridge, MA: MIT Press.
- 896 Barberis, S. D. (2013). Functional analyses, mechanistic explanations, and explanatory tradeoffs. *Journal*
897 *of Cognitive Science*, 14(3), 229–251.
- 898 Barrett, D. (2014). Functional analysis and mechanistic explanation. *Synthese*. doi:10.1007/
899 s11229-014-0410-9.
- 900 Batterman, R. (2013). The tyranny of scales. In R. W. Batterman (Ed.), *The Oxford handbook of philosophy*
901 *of physics* (pp. 255–286). New York: oxford University Press.
- 902 Bechtel, W. (2015). Investigating neural representations: The tale of place cells. *Synthese*.
- 903 Bechtel, W. (2001). Cognitive neuroscience: Relating neural mechanisms and cognition. In P. Machamer, P.
904 McLaughlin, & R. Grush (Eds.), *Philosophical reflections on the methods of neuroscience*. Pittsburgh,
905 PA: University of Pittsburgh Press.
- 906 Bechtel, W. (2008). *Mental mechanisms: Philosophical Perspectives on cognitive neuroscience*. London:
907 Routledge.
- 908 Bechtel, W., & Abrahamsen, A. (2013). Thinking dynamically about biological mechanisms: Networks of
909 coupled oscillators. *Foundations of Science*, 18, 707–723.
- 910 Bechtel, W., & Shagrir, O. (forthcoming). The non-redundant contributions of Marr’s three levels of analysis
911 for explaining information processing mechanisms. *Topics in Cognitive Science*.
- 912 Bickle, J. (2003). *Philosophy and neuroscience: A ruthlessly reductive account*. Dordrecht: Kluwer.
- 913 Bickle, J. (2006). Reducing mind to molecular pathways: Explicating the reductionism implicit in current
914 cellular and molecular neuroscience. *Synthese*, 151, 411–434.
- 915 Bogen, J. (2005). Regularities and causality: Generalizations and causal explanations. *Studies in History*
916 *and Philosophy of Biological and Biomedical Sciences*, 36, 397–420.
- 917 Boone, W., & Piccinini, G. (unpublished). Mechanistic abstraction.
- 918 Burge, T. (2010). *Origins of objectivity*. Oxford: Oxford University Press.
- 919 Caianiello, E. R. (1961). Outline of a theory of thought processes and thinking machines. *Journal of*
920 *Theoretical Biology*, 1(2), 204–235.
- 921 Chemero, A., & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science.
922 *Philosophy of Science*, 75, 1–27.
- 923 Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computa-
924 tional explanation in neuroscience. *Synthese*, 191(2), 127–154.
- 925 Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *Journal of Philosophy*,
926 78, 67–90.
- 927 Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain*. Cambridge, MA:
928 MIT Press.
- 929 Craver, C. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford
930 University Press.
- 931 Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA: MIT Press.
- 932 Cummins, R. (2000). ‘How does it work?’ vs. ‘What are the laws?’ Two conceptions of psychological expla-
933 nation. In K. F. C. & W. R. A. (Eds.), *Explanation and cognition*. Cambridge: Cambridge University
934 Press.
- 935 Daugman, J. G. (1990). Brain metaphor and brain theory. In E. L. Schwartz (Ed.), *Computational neuro-*
936 *science* (pp. 9–18). Cambridge, MA: MIT Press.
- 937 Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling*
938 *of neural systems*. Cambridge, MA: MIT Press.
- 939 Dennett, D. C. (1978). *Brainstorms*. Cambridge, MA: MIT Press.
- 940 Dretske, F. I. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.
- 941 Dretske, F. I. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- 942 Egan, F., & Matthews, R. (2006). Doing cognitive neuroscience: A third way. *Synthese*, 153, 377–391.
- 943 Ermentrout, G. B., & Terman, D. H. (2010). *Mathematical foundations of neuroscience*. New York: Springer.

- 944 Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford: Oxford
 945 University Press.
- 946 Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 6,
 947 205–254.
- 948 Fodor, J. A. (1968a). *Psychological explanation*. New York: Random House.
- 949 Fodor, J. A. (1968b). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy*,
 950 65, 627–640.
- 951 Fodor, J. A. (1974). Special sciences. *Synthese*, 28, 77–115.
- 952 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.
- 953
- 954 Fodor, J. A. (1998). *Concepts*. Oxford: Clarendon Press.
- 955 Fox, P. T., Minton, M. A., Raichle, M. E., Miezin, F. M., Allman, J. M., & Van Essen, D. C. (1986). Mapping
 956 human visual cortex with positron emission tomography. *Nature*, 323, 806–809.
- 957 Fresco, N. (2014). *Physical computation and cognitive science*. New York: Springer.
- 958 Freud, S. (1895/1966). Project for a scientific psychology. In E. Jones (Ed.) & J. Strachey (Trans.), *The
 959 standard edition of the complete psychological works of Sigmund Freud* (Vol. 1, pp. 295–397). London:
 960 Hogarth Press.
- 961 Gallistel, R. G., & King, A. P. (2009). *Memory and the computational brain: Why cognitive science will
 962 transform neuroscience*. New York: Wiley/Blackwell.
- 963 Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(3), S342–S353.
- 964 Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application
 965 to conduction and excitation in nerve. *Journal of Physiology*, 117, 500–544.
- 966 Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*.
 967 New York: Oxford University Press.
- 968 James, W. (1890/1983). *The principles of psychology*. Cambridge, MA: Harvard University Press.
- 969 Jilk, D., Lebiere, C., O'Reilly, R., & Anderson, J. (2008). SAL: An explicitly pluralistic cognitive archite-
 970 cture. *Journal of Experimental and Theoretical Artificial Intelligence*, 20(3), 197–218.
- 971 Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and con-
 972 sciousness*. New York: Cambridge University Press.
- 973 Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, 183(3), 339–
 974 373.
- 975 Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical models. *Philosophy of Science*,
 976 78(4), 601–627.
- 977 Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *Journal of General Physiology*,
 978 59, 734–766.
- 979 Kosslyn, S. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- 980 Kosslyn, S. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.
- 981 Kosslyn, S., & Van Kleck, M. H. (1990). Broken brains and normal minds: Why humpty–dumpty needs a
 982 skeleton. In E. L. Schwartz (Ed.), *Computational neuroscience*. Cambridge, MA: MIT Press.
- 983 Kosslyn, S., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. New York: Oxford
 984 University Press.
- 985 Laird, J. E. (2012). *The soar cognitive architecture*. Cambridge, MA: MIT Press.
- 986 Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial
 987 Intelligence*, 33, 1–64.
- 988 Levy, A. (2013). What was Hodgkin and Huxley's achievement? *British Journal for Philosophy of Science*.
 989 doi:[10.1093/bjps/axs043](https://doi.org/10.1093/bjps/axs043).
- 990 Levy, A., & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science*,
 991 80(2), 241–261.
- 992 Lycan, W. (1981). Form, function, and feel. *Journal of Philosophy*, 78, 24–50.
- 993 Lycan, W. (1990). The continuity of levels of nature. In W. Lycan (Ed.), *Mind and cognition*. Malden, MA:
 994 Blackwell.
- 995 Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1),
 996 1–25.
- 997 Maley, C., & Piccinini, G. (2013). Get the latest upgrade: Functionalism 6.3.1. *Philosophia Scientiae*, 17(2),
 998 135–149.
- 999 Marr, D. (1982). *Vision*. San Francisco: W. H. Freeman and Company.

- 1000 McClelland, J. L., & Lambon Ralph, M. A. (Eds.). (2013). *Cognitive neuroscience: Emergence of mind*
 1001 *from brain. The biomedical & life sciences collection*. London: Henry Stewart Talks Ltd.
- 1002 McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin*
 1003 *of Mathematical Biophysics*, 5, 115–133.
- 1004 Milkowski, M. (2013). *Explaining the computational mind*. Cambridge, MA: MIT Press.
- 1005 Morgan, A. (2014). Representations gone mental. *Synthese*, 191(2), 213–244.
- 1006 Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4, 135–183.
- 1007 Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.
- 1008 Newell, A., & Simon, H. A. (1976). Computer science as an empirical enquiry: Symbols and search.
 1009 *Communications of the ACM*, 19, 113–126.
- 1010 O’Doherty, J., Hampton, A., & Kim, H. (2007). Model-based fMRI and its application to reward learning
 1011 and decision making. *Annals of the New York Academy of Sciences*, 1104, 35–53.
- 1012 O’Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Under-*
 1013 *standing the mind by simulating the brain*. Cambridge, MA: MIT Press.
- 1014 O’Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors. (2014). *Computational cognitive*
 1015 *neuroscience*. Wiki Book (2nd ed.). <http://ccnbook.colorado.edu>.
- 1016 Piccinini, G. (unpublished). Activities are manifestations of causal powers.
- 1017 Piccinini, G. (2007). Computational modeling vs. computational explanation: Is everything a turing machine,
 1018 and does it matter to the philosophy of mind? *Australasian Journal of Philosophy*, 85(1), 93–115.
- 1019 Piccinini, G. (2008). Computers. *Pacific Philosophical Quarterly*, 89(1), 32–73.
- 1020 Piccinini, G., & Bahar, S. (2013). Neural computation and the computational theory of cognition. *Cognitive*
 1021 *Science*, 34, 453–488.
- 1022 Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mech-
 1023 *anism sketches*. *Synthese*, 183(3), 283–311.
- 1024 Piccinini, G., & Maley, C. (2014). The metaphysics of mind and the multiple sources of multiple realizability.
 1025 In M. Sprevak & J. Kallestrup (Eds.), *New waves in the philosophy of mind*. New York: Palgrave
 1026 Macmillan.
- 1027 Piccinini, G., & Scarantino, A. (2011). Information processing, computation, and cognition. *Journal of*
 1028 *Biological Physics*, 37(1), 1–38.
- 1029 Posner, M. I. (1976). *Chronometric explorations of mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- 1030 Posner, M. I., & Raichle, M. E. (1994). *Images of mind*. New York: Scientific American Books.
- 1031 Povich, M. (forthcoming). Mechanisms and model-based fMRI. *Philosophy of Science*.
- 1032 Putnam, H. (1967). *Psychological predicates. Art, philosophy, and religion*. Pittsburgh, PA: University of
 1033 Pittsburgh Press.
- 1034 Putnam, H. (1975). Philosophy and our mental life. In H. Putnam (Ed.), *Mind, language and reality:*
 1035 *Philosophical papers* (Vol. 2, pp. 291–303). Cambridge: Cambridge University Press.
- 1036 Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review*,
 1037 88, 16–45.
- 1038 Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- 1039 Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*, 25(2),
 1040 157–237.
- 1041 Pylyshyn, Z. W. (2003). Return of the mental image: Are there really pictures in the head? *Trends in*
 1042 *Cognitive Science*, 7(3), 113–118.
- 1043 Ramsey, W. M. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.
- 1044 Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*.
 1045 Washington, DC: Spartan.
- 1046 Roskies, A. (2009). Brain–mind and structure–function relationships: A methodological response to Colt-
 1047 *heart*. *Philosophy of Science*, 76(5), 927–939.
- 1048 Rumelhart, D. E., McClelland, J. M., & The PDP Research Group. (1986). *Parallel distributed processing:*
 1049 *Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- 1050 Rusanen, A.-M., & Lappi, O. (2007). The limits of mechanistic explanation in cognitive science. In S. Vos-
 1051 *niadou, D. Kayser, & A. Protopapas (Eds.), Proceedings of the European cognitive science conference*
 1052 *2007* (pp. 284–289). Howe: Lawrence Erlbaum Associates.
- 1053 Scarantino, A. (2015). Information as a probabilistic difference maker. *Australian Journal of Philosophy*.
 1054 doi:10.1080/00048402.2014.993665.
- 1055 Schaffner, K. F. (2008). Theories, models, and equations in biology: The heuristic search for emergent
 1056 simplifications in neurobiology. *Philosophy of Science*, 75, 1008–1021.

- 1057 Shagrir, O. (2010a). Brains as analog-model computers. *Studies in History and Philosophy of Science*, 41,
1058 271–279.
- 1059 Shagrir, O. (2010b). Marr on computational-level theories. *Philosophy of Science*, 77, 477–500.
- 1060 Smart, J. J. C. (1959). Sensations and brain processes. *The Philosophical Review*, 68(2), 141–156.
- 1061 Smart, J. J. C. (2007). The mind/brain identity theory. In E. N. Zalta (Ed.), *The Stanford encyclopedia of*
1062 *philosophy* (Summer 2007 ed.). <http://plato.stanford.edu/archives/sum2007/entries/mind-identity/>.
- 1063 Stein, R. (1965). A theoretical analysis of neuronal variability. *Biophysical Journal*, 5(2), 173–194.
- 1064 Thagard, P. (2007). Coherence, truth, and the development of scientific knowledge. *Philosophy of Science*,
1065 74, 28–47.
- 1066 Treisman, A. (1996). The binding problem. *Current Opinion in Neurobiology*, 6(2), 171–178.
- 1067 Treisman, A. (2009). Attention: Theoretical and psychological perspectives. In M. Gazzaniga (Ed.), *The*
1068 *cognitive neurosciences* (4th ed., pp. 189–204). Cambridge, MA: MIT Press.
- 1069 Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12,
1070 97–136.
- 1071 Van Eck, D., & Weber, E. (2014). Function ascription and explanation: Elaborating an explanatory utility
1072 desideratum for ascriptions of technical functions. *Erkenntnis*. doi:10.1007/s10670-014-9605-1.
- 1073 Vartanian, A. (1973). In P. P. Wiener (Ed.), *Dictionary of the history of ideas: Studies of selected pivotal*
1074 *ideas*. New York: Scribners.
- 1075 Waxman, S. (1972). Regional differentiation of the axon: A review with special reference to the concept of
1076 the multiplex neuron. *Brain Research*, 47, 269–288.
- 1077 Weber, M. (2005). *Philosophy of experimental biology*. Cambridge: Cambridge University Press.
- 1078 Weber, M. (2008). Causes without mechanisms: Experimental regularities, physical laws, and neuroscien-
1079 tific explanation. *Philosophy of Science*, 75(5), 995–1007.
- 1080 Weisberg, M. (2013). *Simulation and similarity: Using models to understand the world*. Oxford: Oxford
1081 University Press.
- 1082 Weiskopf, D. (2011). Models and mechanisms in psychological explanation. *Synthese*, 183(3), 313–338.
- 1083 Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of
1084 model neurons. *Biophysical Journal*, 12, 1–24.
- 1085 Winsberg, E. (2010). *Science in the age of computer simulation*. Chicago: University of Chicago Press.