12-14-2016

Reconciling New Mechanism and Psychological Explanation: A Pragmatic Approach

Michael De Vivo
Georgia State University

RECONCILING NEW MECHANISM AND PSYCHOLOGICAL EXPLANATION:

A PRAGMATIC APPROACH

by

MICHAEL J. DE VIVO

Under the Direction of Andrea Scarantino, PhD

ABSTRACT

Recently, Gualtiero Piccinini and Carl Craver (2011) have argued that functional analyses in psychology lack explanatory autonomy from explanations in neuroscience. In this thesis I argue against this claim by motivating and defending a pragmatic-epistemic conception of autonomous psychological explanation. I argue that this conception of autonomy need not require that functional analyses be distinct in kind from neural-mechanistic explanations. I use the framework of Bas van Fraassen's Pragmatic Theory of Explanation (van Fraassen 1980) to show that explanations in psychology and neuroscience can be seen as seeking understanding of autonomous levels of mechanistic phenomena.

INDEX WORDS: Mechanisms, Explanation, Autonomy, Psychology, Neuroscience, Pragmatics

RECONCILING NEW MECHANISM AND PSYCHOLOGICAL EXPLANATION:

A PRAGMATIC APPROACH

by

MICHAEL J. DE VIVO

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2016

RECONCILING NEW MECHANISM AND PSYCHOLOGICAL EXPLANATION:

A PRAGMATIC APPROACH

by

MICHAEL J. DE VIVO

Committee Chair: Andrea Scarantino

Committee: Daniel Weiskopf

Neil Van Leeuwen

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2016

ACKNOWLEDGEMENTS

I would like to thank Andrea Scarantino for his constant intellectual criticism, encouragement, and patience. As an advisor, Andrea helped me to become a more careful and rigorous philosopher. As a mentor, he helped me overcome some anxiety over doing philosophy. Second, thanks to my committee members, Dan Weiskopf and Neil Van Leeuwen, for feedback on earlier drafts as well as the intellectual growth I experienced in their courses. Third, thanks to my GSU friends for their philosophical stimulation and personal support. Thanks especially to Amanda Haskell for reading every draft of this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

The ‘new mechanistic’ philosophy of science holds that many sciences aim to discover and describe *mechanisms* (Andersen 2014). One crucial question about mechanisms concerns when they are *explanatorily relevant* to a phenomenon. According to some mechanists, “to explain a phenomenon, one must cite mechanistic information, i.e. specify underlying parts and their organization” (Levy 2013: 2, ms.).[1]

Recently, Gualtiero Piccinini and Carl Craver (2011) (‘P&C’ hereafter) have sought to integrate psychological explanation into a mechanistic framework.[2] Specifically, P&C argue that *functional analysis* in psychology is a kind of mechanistic explanation, and crucially, because of this, psychology is not explanatorily *autonomous from* neuroscience.[3] P&C write:

> [W]e argue that a complete […] explanation of a phenomenon in terms of functional properties must respect constraints imposed by the structures that possess those functional properties—that is, it requires fitting the functional properties within a mechanism. (2011: 286)

P&C’s arguments target the so-called ‘Received View’, a central tenet of which is that psychology is *explanatorily autonomous* from neuroscience.[4] At first blush, this notion can be introduced roughly as follows:

> To say that psychology has *explanatory autonomy* is to say that cognitive models are sufficient by themselves to give adequate explanations of various psychological phenomena. (Weiskopf forthcoming: 13, manuscript, original italics)

> A scientific explanation can be regarded as *explanatorily autonomous* if it can provide genuine explanatory statements, which appeal only to the taxonomic categories used within that discipline. (Feest 2003: 938, original italics)

[1] This view has gained favor in biology (e.g. Bechtel and Richardson 1993) and neuroscience (e.g. Craver 2007a).

[2] All references to P&C hereafter are to their (2011).

[3] The notions of functional analysis and mechanistic explanation are clarified in the next section.

[4] They cite Jerry Fodor (1968) and Robert Cummins (1975, 1983) as early proponents of this view. The Received View has by no means fallen out of favor in contemporary circles. Recent defenders include David Barrett (2014) and Lawrence Shapiro (2016). Other proponents arguably include Frances Egan (forthcoming), Uljana Feest (2003), and Daniel Weiskopf (2011, forthcoming).

To a first approximation, then, explanatory autonomy is psychology's ability to explain its target phenomena with proprietary vocabulary and models. In this sense, psychology gives genuine explanations on its own theoretical and taxonomic grounds, and need not appeal to other disciplines (most relevantly neuroscience) to validate its explanations.[5]

According to P&C, autonomous psychological explanation requires that psychological explanation and mechanistic explanation differ in kind, a view they call *distinctness*. They write that "if functional analysis is *a kind of* mechanistic explanation, as we argue, then functional analysis cannot be *autonomous from* mechanistic explanations" (pp. 287-288). In light of P&C's claims, defenders of autonomy ('Autonomists' hereafter) must ask whether autonomy really depends on distinctness after all. If it does, as some recent Autonomists seem to suggest (e.g. Shapiro 2016, Weiskopf forthcoming), then defending distinctness should be a primary objective for Autonomists. However, if autonomy does *not* depend on distinctness, a possibility that has gone largely unexplored, then Autonomists have reason to seek a different foundation for autonomy.

Several reasons suggest that distinctness is an untenable position for the Autonomist. First, psychology does pay close attention to neuroscience (Bechtel 2008a; Boone and Piccinini 2016).[6] Second, functional analysis is by no means proprietary to psychology: many mechanistically oriented sciences, most notably biology and neuroscience, rely on functional analysis.[7] Third, many functional analyses are intended as *mechanistic abstractions* (Boone and Piccinini, unpublished ms.) or *mechanism sketches* (P&C), differing only in degree, but not in kind, from full-blown

[5] As will become clear below, this does not mean that psychological explanations are not confirmed, disconfirmed, or revised according to data from neuroscience.
[6] For example, Bechtel (2008) discusses extensively how the functional decomposition of memory was largely guided by discoveries of different hippocampal structures.
[7] Similarly, Piccinini (2006) and Kaplan (2011) argue that functional explanation, specifically *computational* explanation, is a legitimate form of explanation in computational neuroscience, and thus not proprietary to psychology.

mechanistic explanations. Finally, distinctness leaves explanations in psychology and neuroscience *epistemically fragmented*: it is hard to see how explanations in these disciplines can be woven into a unified body of scientific knowledge if they differ in kind.

P&C's assault on autonomy is primarily motivated by this latter concern of explanatory unification.[8] Specifically, they write the following:

> Our argument against the autonomy thesis […] leads to a new understanding of how psychology and neuroscience should be integrated—explanatory unification will be achieved through the integration of findings from different areas of neuroscience and psychology into a description of multilevel mechanisms. (P&C: 284-285)

P&C view autonomy as *the* main obstacle to explanatory unification. Distinctness, for them, is autonomy's life source, and hence must be cut off to achieve explanatory unification. However, I think that P&C strongly overestimate the undermining force of autonomy on explanatory unification. This fact is, I think, overshadowed by their assumption that distinctness is necessary for autonomy, an assumption of many Autonomists themselves. But if autonomy can be defended without distinctness, it is much less clear that Autonomy hinders explanatory unification.

With few exceptions (e.g. Craver 2007a; Hardcastle 1996; Sirtes 2010), discussions of mechanisms have said little about the *pragmatic* dimensions of mechanistic explanation. Even fewer have attempted to ground mechanistic explanation in a pragmatic *theory* of explanation (e.g. Sirtes 2010). However, I believe that overlooking the pragmatics of explanation has led P&C to see autonomy as stifling the integration of psychology and neuroscience when it poses no such threat.

In this thesis, I defend a pragmatic approach to autonomy which aims to reconcile functional analysis and mechanistic explanation. My thesis comprises four interrelated claims. First, autonomy is ineluctably tied to the *pragmatics of explanation* in the sense that autonomy depends on

---

[8] The title of their paper, "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches," highlights this concern.

the extent to which explanations can answer different *why-questions*. Second, the sense of autonomy most relevant to explanatory practice in psychology is bound tightly, though not exclusively, to desiderata emphasized by pragmatic theories of explanation (e.g. Faye 2007; van Fraassen 1980). Third, rejections of autonomous psychological explanation overlook such pragmatic considerations. Finally, a pragmatic defense of autonomy avoids the mechanist charge that autonomous psychological explanation hinders explanatory integration of psychology and neuroscience.

This thesis runs as follows: In §2, I compare and contrast functional analysis and mechanistic explanation. In §3 I argue that P&C's target conception of autonomy presupposes an *ontic* mode of 'explanation', which excludes crucial pragmatic desiderata of autonomy.[9] In §4 I argue that the autonomy that best reflects the actual concerns of psychologists is pragmatic and epistemic in nature. In §5 I couch autonomy in a pragmatic explanatory-theoretic framework (e.g. van Fraassen 1980). In §6 I argue that autonomy is compatible with a mechanistic framework of psychological explanation. I conclude in §7 that, contra P&C, autonomy is no obstacle to explanatorily integrating psychology and neuroscience.

2 FUNCTIONAL ANALYSIS AND MECHANISTIC EXPLANATION[10]

First, I use 'mechanism' along the lines of the following definition:

> A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel and Abrahamsen 2005: 423)

[9] See Craver (2014), Illari (2013), and Wright (2012) for helpful discussions of different senses of 'explanation'.

[10] Most of these terms and concepts are philosophically contentious: philosophers disagree over what these terms and concepts amount to, or at least they tend to use them in different ways. Thus, my characterizations of the following will stick closely to P&C's usage.

On this definition, mechanisms are composites of parts, causal activities (or operations), and arrangements. They produce the phenomena they do *because* of the component, causal, and organizational properties they have. Accordingly, *mechanistic explanations* describe component, causal, and organizational properties of mechanisms.[11] Additionally, mechanisms are 'multilevel' in the sense that whole mechanisms have mechanisms as parts. Each level consists of parts, causal activities, and structural arrangements (Bechtel 2008a).

Second, 'function' is synonymous with 'what something does'. Thus the functions of encoding, representing, and storing are individuated by their typical effects within a larger system (e.g. the working memory system). Functions typically make no reference to the physical properties that realize them: suitably organized silicon chips or neural pathways can realize an encoder. *Functional analysis*, the strategy by which phenomena are explained in functional terms, is understood here in Cummins' sense:

> Functional analysis consists in analyzing a disposition into a number of less problematic dispositions such that the programmed manifestation of these analyzing dispositions amounts to a manifestation of the analyzed disposition. (Cummins 1983: 28)

A disposition is a tendency of something to function in a regular way. Normally developed humans have the disposition to transform 2-D retinal images into 3-D representations. *Programmed* dispositions are ones arranged such that they endow some larger system with its own dispositions.[12] Functional analyses aim to study the relationship between programmed dispositions and the dispositions of the systems that contain them.

---

[11] For example, long-term potentiation in the hippocampus is explained by NMDA and GABA (component parts), chemical binding (operations), and multitudinous pathways and projection sites (organizations).

[12] To use Cummins' (1983) example, a car assembly line divides simple tasks such as attaching wheels and fastening bolts, and the sequence of such tasks (among others) endows the assembly line as a whole with the disposition to produce cars.

An example of functional analysis is Alan Baddeley's model of working memory (Barrett 2014; Weiskopf forthcoming). This functional analysis attempts to explain working memory by specifying its *functional* architecture, i.e., components of working memory individuated by the effects they have on the operation of working memory as a whole.

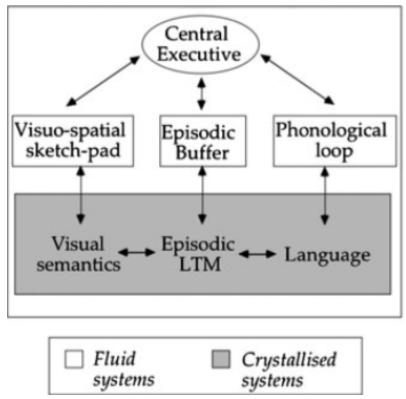

Figure 1 An updated version of the Baddeley's functional analysis of working memory

In Baddeley's model, each box refers to a functional unit that performs a certain operation. For example, the phonological loop is devoted to processing and storing verbal information, and the episodic buffer is devoted to transforming sensory information into a multimodal format (Baddeley 2012). The arrows in the model depict the highly abstract causal structure of this functional architecture. As Weiskopf (forthcoming) points out, this model is spatially neutral: the lengths of the arrows do not indicate the spatial distance between these functional units. In contrast, this is often the case in diagrams of mechanistic phenomena, where size and shape of elements in a diagram are drawn to scale with the size and shape of corresponding elements in the target phenomenon.

Functional analysis and mechanistic explanation compare and contrast in four main ways. First, both attend to parts, operations, and organizations of systems. However, each does so with an eye to distinct kinds of properties (functional versus mechanistic properties). Second, both decompose capacities of interest into *sub*capacities. However, decomposition is done by reference to distinct kinds of properties. Third, while both attend to organizational constraints, mechanistic explanations appeal to *spatial* properties (e.g. size, shape, and location) while functional analyses, being abstract, allegedly do not. Finally, both give bona fide *causal* and *constitutive* explanations, but differ in the organizational level at which causal relations and constitutive relations are judged explanatorily relevant (Shapiro 2016; Weiskopf 2011, forthcoming). A causal relation is when something is a cause or effect of something else. A constitutive relation is being a part of something else or having something else as a part.

3 THE ONTIC AND EPISTEMIC CONCEPTIONS OF 'AUTONOMY'

In this section I argue that P&C's sense of 'autonomy' relies on an ontic conception of 'explanation'. This fact strips their target notion of autonomy of epistemic desiderata that Autonomists value most about autonomy.

3.1 Autonomy as 'lack of direct constraints'?

According to P&C, the relevant sense of explanatory autonomy is 'no-direct-constraints autonomy':

> [A] functional analysis directly constrains a mechanistic explanation if and only if the functional properties described by a functional analysis restrict the range of structural components and component organizations that might exhibit those capacities; a mechanistic explanation directly constraints a functional analysis if and only if the structural components and component organization described by the mechanistic explanation restrict the range of functional properties exhibited by those components thus organized. (P&C 2011: 289)[13]

On this conception, the phenomena studied by psychologists (i.e. functions or functional architectures) and the phenomena studied by neuroscientists (i.e. neural mechanisms or architectures)

[13] Note that this passage says more than just that functional architecture of a system is *realized* by some mechanism. P&C call this an "indirect constraint," which they see as consensus for all views.

place no restrictions on each other. The "range restrictions" P&C speak of involve the degree of *causal* influence between functional properties and structural properties. Causal influence is specifiable on an interventionist (or 'manipulationist') framework (Woodward 2003).[14]

On P&C's view, two properties A and B directly constrain each other's causal powers if and only if interventions on A produces interventions on B, and vice versa.[15]

Given their characterization, I understand P&C as making the following argument:

1. Functional analyses and mechanistic explanations directly constrain each other.
2. Two explanations are explanatorily autonomous if and only if they do not directly constrain each other.

   Therefore,

3. Psychological explanations and mechanistic explanations are not autonomous.

My main issue is with Premise 2: P&C's definition of 'autonomy' that they aim to reject. Once this premise is challenged, I think Premise 1 becomes much harder to swallow.

3.2 Disambiguating two modes of 'explanation'

My issue with Premise 2 is that we cannot infer explanatory constraints from the mere presence of direct property constraints. P&C do not say much about what makes property constraints explanatorily relevant to functional analyses. And it is far from clear how the mere presence of property constraints alone could make an *explanatory* difference to a functional analysis. Given this,

---

[14] P&C do not directly relate their view of autonomy to an interventionist framework, but the emphasis on property constraints and range restrictions seems strongly suggestive of this orientation.

[15] This is the logic behind function localization experiments in neuropsychology. For example, an experimenter might intervene on a brain region using transcranial magnetic stimulation to observe its influence on a specific cognitive function, and subsequently, manipulate that cognitive function (e.g. with a dichotic listening task) to observe its influence on a specific brain region.

how can we know which property constraints make a difference to the phenomena we intend functional analyses to represent? Moreover, in what way does *representing* property constraints advance the epistemic aim of functional analyses to facilitate understanding?

I believe this unclarity arises from an ambiguity at play between two modes of 'explanation' in P&C's discussion: the *ontic* mode and the *epistemic* mode (Craver 2014; Illari 2013; Wright 2012). The ontic mode sees explanations as synonymous with causes, laws, statistical relations, and presumably, property constraints.[16] The epistemic mode, in contrast, ties explanations to our psychological capacities to understand phenomena by representing them in certain ways (Bechtel and Abrahamsen 2005; Bechtel 2008a).[17]

P&C's 'no-direct-constraints autonomy' seems squarely aligned with the ontic mode. For them, whether two explanations are autonomous is a mind-independent fact about property constraints. Recall that, on their view, if two properties are *mutually manipulable*, they directly constrain each other. In contrast, functional analyses—ways of *representing* phenomena in the world—are squarely aligned with the epistemic mode. Their primary purpose is to facilitate our *understanding of* psychological phenomena.[18]

In light of clarifying this ambiguity about 'explanation', we can see why P&C's definition of 'autonomy' is problematic for their rejection of autonomy. Suppose a functional property (say, being an encoder) and a structural property (say, being part of hippocampal region X) directly constrain each other. This is an ontic matter: whether they mutually manipulate each other is a

[16] In this mode, an object's falling to earth is explained by Newton's law of universal gravitation. This mode entails that explanations exist independently of human knowledge or understanding.
[17] In this mode, 'explanation' picks out a *representation* (e.g. a model or diagram) of some phenomenon (Bechtel and Abrahamsen 2005; Bechtel 2008a). The epistemic mode thus sees explanation as fundamentally tied to our *understanding*: we have explained object motion when *we* arrive at an intelligible understanding of Newton's laws.
[18] In the next section, I address the concern about what is meant by the term 'understanding' in the context of explanation.

mind-independent fact. In P&C's view, these properties are thus not autonomous from each other. But it is a separate question whether representing such constraints serves psychologists' pragmatic aim of deepening understanding.

When it comes to figuring out *which* constraints to represent in functional analyses, pragmatic and epistemic concerns trump ontic ones. In this way, P&C's Premise 2 trades on an ambiguity of 'explanation'. Consequently, Premise 1—the claim that functional analyses and mechanistic explanations directly constrain each other—leaves it open for debate whether direct property constraints undermine the autonomy of explanatory *models* per se. To summarize, my argument against P&C is this:

(1) There are two modes of explanation: ontic and epistemic.

(2) One's definition of 'explanatory autonomy' is constrained by which mode of 'explanation' is presupposed in the definition.

(3) If a strictly ontic mode of explanation is presupposed, then this is irrelevant to the explanatory autonomy of functional analyses.

(4) P&C's conception of 'explanatory autonomy' presupposes the ontic mode of 'explanation'.

Therefore,

(5) P&C's conception of 'explanatory autonomy' is irrelevant to the explanatory autonomy of functional analyses from mechanistic explanations.

The upshot of this argument is that the ontic and the epistemic modes of 'explanation' entail two different notions of explanatory autonomy. P&C's failure to acknowledge this distinction leads them to apply conclusions about the ontic sense of 'autonomy' to the epistemic sense.

P&C can challenge any one of these premises, but the most contestable are (2) and (3). Challenging Premise 2, P&C could argue that 'explanatory autonomy' is *theory neutral* between the ontic and epistemic modes. For example, P&C characterize no-direct-constraints autonomy as allowing for interdisciplinary ignorance between psychology and neuroscience:

> [No-direct-constraints autonomy] neatly divides the explanatory labor along traditional disciplinary lines and thus relieves members of each discipline of learning overly much about the other discipline. On one hand, psychologists are given the task of uncovering the functional organization of the mind without worrying about what neuroscientists do. On the other hand, neuroscientists are given the task of discovering neural mechanisms without having to think too hard about how the mind works. (P&C: 290)

This passage seems to present no-direct-constraints autonomy under an epistemic light by allowing disciplines to advance their explanatory pursuits with a considerable degree of ignorance about each other's findings.

In spite of this epistemic gloss, I contend that P&C's rejection of Autonomy succeeds to the extent that the sorts of explanations at issue are ontic in nature, not epistemic. If P&C's argument were directed toward the sense of 'explanation' that puts the autonomy of *functional analysis* at risk (as opposed to the autonomy of the properties that functional analyses *depict*), they would have to shift their target to a different sense of 'autonomy'—one constrained by an epistemic mode of 'explanation'.

This is because which constraints apply to explanatory models is a matter of which constraints are relevant to supporting an understanding of some phenomenon. Once understanding is factored into our explanatory desiderata, the relevant sense of 'explanation' shifts from the ontic mode to the epistemic mode. In this mode, property constraints by themselves do not undermine autonomy: 'autonomy' is a matter of how well a model independently grounds understanding. Absent an account of how the *representation* of property constraints would support understanding, property constraints seem to undermine a different sense of explanatory autonomy from the sense entailed by epistemically-oriented explanations. It is in this way that one's definition of 'explanatory autonomy' is constrained by which mode of 'explanation' is presupposed.

3.3 Confirmatory autonomy versus explanatory autonomy

Now recall my Premise 3: Ontic explanatory autonomy has no direct impact on the explanatory autonomy of functional analyses from mechanistic explanations. P&C might argue that ontic constraints *do* constrain explanatory models, and in this way, property constraints *do* undermine the explanatory autonomy of functional analyses. Thus they write the following:

> [I]f a sub-capacity is a genuinely explanatory part of the whole capacity, as opposed to an arbitrary partition (a mere piece or temporal slice), *it must be exhibited by specific components or specific configurations of components*. In the systems with which psychologists and neuroscientists are concerned, the sub-capacities are not ontologically primitive; they belong to structures and their configurations. The systems have the capacities they have in virtue of their components and organization. (P&C: 293, my emphasis)

P&C seem to be suggesting that functional analyses acquire explanatory power by representing constraints imposed on the mind's functional architecture by the brain's neural architecture.[19] For example, if a functional analysis has a box labeled 'phonological loop', then there must *be* a mechanism in the brain that *functions as* a phonological loop. Hence they write "if the [structural] components predicted by a given [functional] analysis are not there…you must rule out that [functional] analysis in favor of another for the case in question" (p. 293).

The above quote from P&C suggest that functional analyses are *confirmable* by whether or not the brain actually instantiates the functional architectures predicted by those analyses. However, this poses no threat to my Premise 3, which says only that ontic constraints have no direct influence on the *explanatory* autonomy of functional analyses. Premise 3 does not say that ontic constraints make no difference to the *correctness* or *revisability* of a functional analysis.

Indeed, I agree with P&C that a functional analysis ought to reflect relevant features of the instantiating mechanism. This seems to be the sense in which P&C urge psychologists to "let

---

[19] It is this sense in which they think that the division of working memory into stages of encoding, storage, and retrieval "places direct constrains on components, their functions, and their organization" (p. 293)

knowledge of neural mechanisms constrain their hypotheses" (p. 285). But looking to neuroscience to *confirm* psychology's functional analyses is different from looking to neuroscience aid psychology's explanatory work.

Granted, conceding to P&C that psychologists must pay attention to neuroscientific findings does undermine a certain *sense* of autonomy. However, as Roth and Cummins (forthcoming) have rightfully noted, the sense that is undermined is merely *confirmatory autonomy*:

> Knowledge of neural structures is undoubtedly relevant to settling the question of which [functional] analysis is correct, but bringing such knowledge to bear in this instance would be an exercise in confirming a proposed analysis, not explaining a capacity. (p. 18, ms.)

Functional analyses in psychology are offered as explanations of a particular kind of phenomena: cognitive capacities. Their ability to adequately explain depends on how well they support an understanding of cognitive capacities *as such*, i.e., in terms of the role they play within a *cognitive* system. The ontic constraints that serve to confirm or disconfirm functional analyses do not undermine their explanatory autonomy.

To summarize, I have argued that P&C's characterization of autonomy confounds two modes of 'explanation'. Consequently, the definitional premise of their argument—Premise 2—leaves the relevant sense of 'autonomy' unclarified. I suggested that these two modes of explanation—the ontic and the epistemic—entail distinct senses of 'autonomy', and that P&C's arguments speak only to the ontic sense. However, P&C's claims about the lack of autonomy of functional properties from mechanistic ones do not justify their claim, in Premise 1, that functional analyses and mechanistic explanations, *qua explanatory representations*, directly constrain each other.

4 PRAGMATIC AND EPISTEMIC DESIDERATA OF AUTONOMY

One contentious aspect of P&C's ontic, no-direct-constraints characterization of autonomy is how far removed it is from the predominantly *pragmatic* and *epistemic* concerns of Autonomists. These concerns are expressed clearly in recent discussions of the Received View. Here are just three examples that make this explicit:

> [W]hy not interpret autonomy in the other way P&C suggest, that is, as meaning 'that those who are engaged in functional analysis need *not know* or *pay attention to* what mechanisms are present in the system' (P&C [2011], p. 289). From this perspective, it remains an open question whether functional explanation is autonomous. (Shapiro 2016: 10, my emphasis)

> Explanatory autonomy…is simply the ability to formulate genuine explanations *without knowing* anything about other, even perhaps related, sciences. (Barrett 2014: 2696, fn.1, my emphasis)

> [A]utonomy: the work of explanatory inquiry is divided among many domains, each of which is not merely permitted but required to *black-box* the explanatory models generated by other domains. (Strevens forthcoming: 21, my emphasis)

The moral of these passages is that, given one's explanatory interests, a certain degree of *ignorance* between disciplines is permissible. Moreover, as suggested in the quote from Strevens, ignorance is crucial for the ability of many explanations to support understanding. In similar spirit, I wish to make the case that *most* of what Autonomists value about explanatory autonomy is tied to pragmatic and epistemic concerns, not ontic ones.

4.1 Autonomy: Ignorance, not lack of knowledge

It is important to note the difference between two kinds of epistemic locutions: (1) 'not *knowing* anything about X when explaining Y' and (2) '*ignoring* X when explaining Y'. Discussions of autonomy tend to overlook this difference. P&C seem guilty of this by equating 'not knowing' with 'not paying attention to' in the passage above from Shapiro. However, not knowing something is a much worse epistemic situation to be in than merely ignoring or not paying attention to that thing. For example, lacking knowledge about the hippocampal structures involved in working memory can prevent one from confirming or revising their functional analysis of working memory. To the contrary, ignoring such information—which presupposes that this information is known in the first place—can pay epistemic dividends. In the course of constructing functional analyses, ignoring details about neural implementation serves to highlight details that are of more immediate explanatory interest to investigators. But more crucially, ignorance is also necessary for not being led astray from one's explanatory aims.

Given what P&C say about the epistemological dimension of autonomy, it would seem that autonomy demands that psychologists be able to explain cognitive capacities without *knowing* anything about neuroscience rather than merely *ignoring* known details from neuroscience. This would be a mistake. Abstractions and idealizations—two ways in which scientific explanations ignore or omit some details over others—serve many important pragmatic and epistemic roles.

Boone and Piccinini (unpublished ms.) discuss three such roles. First, an abstraction can highlight as yet unknown causes of a phenomenon, which can guide further inquiry. Second, an abstraction can make a model more "tractable" by virtue of its simplicity and idealization.[20] Finally, abstractions answer to the direct explanatory interests of those investigating some phenomenon. They sum these points up nicely as follows:

> [O]ur epistemic interests and limitations often demand abstractions, namely, abstraction from unknown details, abstraction from some levels of a mechanism in favor of one or more levels, abstractions in the service of mathematical and computational tractability, and abstraction from one or more aspects of one level in favor of other aspects at that level. (Boone and Piccinini, unpublished ms.: 6).

These considerations, I think, highlight the pragmatic and epistemic role of ignorance. At the same time, they highlight why an Autonomist need not be committed to thinking that functional analyses

[20] As Boone and Piccinini note, such idealizations are still epistemically useful for purposes of prediction and discovery (p. 5, ms.).

need be explanatory without knowing anything about neuroscience. To the contrary, *knowing* facts about neuroscience could be an important first step in the psychologist's pragmatic decision of what is and is not permissible to ignore in modeling a cognitive capacity.

This is why, in discussions of autonomy, the difference between not knowing, on the one hand, and ignoring, on the other, is crucial for moving the debate forward. If the standard for autonomy is not needing to know anything about neuroscience in order to develop and revise adequate functional analyses, then it seems unattainable. But if the standard is being permitted to ignore details from neuroscience, then this seems not only attainable, but is already the status quo in psychology (Strevens forthcoming).[21]

4.2 Autonomous explanation and understanding

By 'understanding' I mean understanding *why* some phenomenon occurs. Crucially, while understanding is an epistemic notion *par excellence*, it can be easily wedded to ontic concerns about explanations, such as property constraints, causes, or constitutive relations. Given this, I endorse Michael Strevens' 'Simple View' of understanding (2013, forthcoming), according to which "Understanding why is a matter of grasping facts about the world out there—it is a matter of grasping, roughly, the causes of the phenomenon to be explained" (forthcoming: 6, ms.). On this view, understanding requires one to distinguish relevant from irrelevant causal difference-makers of a phenomenon of interest, and to grasp the reasons for this difference (p. 6).

Let me now more explicitly state my view: Autonomy, in the most relevant sense for psychology, is the unique epistemic contribution that an explanation makes to understanding a phenomenon, such that the absence of this contribution *in particular* would leave a knowledge-gap about that phenomenon.

---

[21] Strevens also makes this point about other special sciences, which I ignore for purposes of my discussion.

For example, sometimes psychologists seek explanations that further clarify the function of a capacity that has already been functionally analyzed. Some such inquiries require appeals to *behavioral* data to intelligibly investigate and better understand the capacity in question. This was the case for better understanding the function of the phonological loop, a component of working memory responsible for temporarily storing verbal information. Baddeley and colleagues studied a patient, PV, who displayed an abnormally low auditory digit span (two items). They used behavioral data about PV's deficit in order to test the hypothesis that the phonological loop governs long-term phonological learning. This was confirmed by PV's inability to learn new word pairs from a foreign language (Russian) while being able to learn word pairs from her native language (Italian).[22] Behavioral data allowed Baddeley and colleagues to make useful inferences about a functional component of working memory—inferences which led to a better understanding of the role of the phonological loop in long-term verbal memory.

4.3 The role of psychologizing in autonomous explanations

At this stage, one might worry that my pragmatic-epistemic conception of autonomy collapses into a kind of *psychologism*: it ties explanations too close to our interests and epistemic limitations (cf. Waskan 2011).[23] In this way, the objection continues, my characterization runs the risk of weakening the interests of those who wish to defend autonomous psychological explanation in the first place. Psychologists presumably want the autonomy of their explanations to be reflected in the ontic structure of the world itself, not just in our understanding of the world.

In response, I think that a certain degree of psychologism is indispensable to science's explanatory practices, specifically to its modeling practices. Explanatory models in science often serve

---

[22] This is recounted in Baddeley (2012).

[23] As Waskan (2011) notes, many defenders of ontic theories of explanation object to epistemic theories by pointing out the risk of psychologism, i.e., "classifying faulty explanations as good and good explanations as faulty" (p. 5, ms.).

the express purpose of helping us draw *intelligible* connections between ontic structures (what P&C and other mechanists are calling 'explanations') and the phenomena they exhibit (Bechtel 2008a; Waskan 2011). Psychologism enters in precisely where science attempts to connect explanations in the ontic sense (i.e. mechanisms themselves) to explanations in the epistemic sense (i.e. models of mechanisms). My appeal to Strevens' Simple View of understanding was intended as one way of reconciling an explanation's intelligibility and its correctness: *explanatory models serve to support understanding by making facts about the world intelligible to us* (Illari 2014; Waskan 2011; Wright 2012).[24] Some degree of psychologism is not inimical to Autonomists' desire to ground autonomy in ontic features of the world.

5 A PRAGMATIC-THEORETIC FRAMEWORK OF AUTONOMY

So far I have argued that autonomy is best seen as a pragmatic and epistemic notion. Further, I have argued that P&C's ontic, no-direct-constraints conception of autonomy overlooks important aspects of autonomy that Autonomists seem to prioritize. My goal in this section is to ground autonomy in a certain explanatory *framework*: The Pragmatic Theory of Explanation ('PTE') (e.g. van Fraassen 1980). In doing so, I hope to clarify the way in which the epistemological concerns of psychology and neuroscience are rooted in distinct pragmatic concerns.

[24] As Illari (2014) points out, those who endorse epistemic explanation (e.g. Bechtel 2008a; Wright 2012) tend to use 'understanding' as a *success* term: we understand a mechanism when our representation of (or text describing) that mechanism advances our knowledge about how that mechanism actually works.

5.1 The pragmatic theory of explanation[25]

First let me introduce PTE. Bas van Fraassen (1980) argued that explanations are answers to why-questions seeking certain kinds of information.[26] The demand for an explanation is always stated against the backdrop of presuppositions stating what the why-question is about (its *topic*), which alternative questions the why-question is contrasted with (its *contrast-class*), and what particular information the question is requesting (its *relevance relation*).

For example, if I ask why Mary laughed at the joke, the topic of my question is the proposition *that Mary laughed at the joke*. The contrast-class could be why *Mary* laughed at the joke (rather than Sally), why Mary *laughed* at the joke (rather than sneered), or why Mary laughed at the *joke* (rather than the speech). And the relevance relation is Mary's *reasons* for laughing at the joke rather than what causally related events occurred in her nervous system to issue in the behavior.

Following van Fraassen, we should think of an explanatory request as a why-question (Q) whose contents are pragmatically determined by their topics (P_k), contrast-classes (X) and relevance relations (R). As a first step to sketching his model, he presents the formal representation of this triad is as follows:

$$Q = \langle P_k, X, R \rangle$$[27]

For ease of reference, let us call Q an *explanatory context*. As a second step, van Fraassen says that a proposition A is relevant to an explanatory context when it bears the relevance relation R to the pair $\langle P_k, X \rangle$ (p. 144). For example, the proposition 'Mary found the joke funny' is relevant to the above scenario involving Mary because Mary finding the joke funny is a possible *reason* for

[25] The term 'pragmatic,' as typically used by PTEs, pertains to psychological facts about human interests, agendas, and epistemic limitations (Woodward 2014).

[26] I restrict attention to van Fraassen's version of PTE because, in my view, it offers the most systematic framework for pinpointing autonomous explanations than other versions of the theory.

[27] This is van Fraassen's own formal apparatus (1980: 144).

her having laughed at (rather than sneer) the joke. As a final step, van Fraassen tells us the form that an answer (*) to a why-question W in some explanatory context must take to count as a direct answer to W:

(*) P_k in contrast to (the rest of) X because A[28]

where, again, P_k is what the question is inquiring *about*, X is a class of propositions that the topic (*) is contrasted with, and A is the reason that (*) is true instead of X. For example:

(*) Mary laughed at the joke (rather than sneered, cried, etc.) because she found it funny.

As van Fraassen notes, answer (*) is a proposition. Furthermore, (*)'s content is determined by the explanatory context, i.e., the same context that determines the content of why-question W (p. 144). Finally, it is worth emphasizing the importance of contrast in (*). The proposition following 'because' gives us two bits of information: it explains (1) why some phenomenon occurred and (2) why some other phenomena {X} *did not* occur instead. The role of contrast thus serves the purpose of drawing a stronger connection between the explanandum (e.g. Mary's laughing) and the explanans (finding something funny) by *ruling out* alternative possibilities.

It is worth drawing some connections between Strevens' Simple View of understanding and van Fraassen's pragmatic framework. First, any answer (*) to a why-question W serves the pragmatic role of facilitating an understanding of the phenomenon that prompted W in context Q. Second, given its role in facilitating understanding, an answer (*) to W must enable Q's audience to distinguish relevant from irrelevant causal difference-makers of the phenomenon, and to grasp reasons for this difference. Thus (*) provides a contrast-class that specifies why a phenomenon is linked to a particular set of causal factors but not others. Furthermore, to facilitate understanding,

[28] Again, this is van Fraassen's formalism (1980: 144).

an answer (*) to a why-question must be intelligibly connected to the relevance relation R presupposed in context Q, such that the information carried by (*) is relayed to the explanatory audience in the intended way.

5.2 Pragmatic explanation as a metric of explanatory autonomy

By pragmatically individuating why-questions according to their explanatory contexts, we can begin to see that explanations address rather parochial knowledge-gaps that appropriate understandings are meant to close. This makes it easier to tell if any given explanation directly addresses the knowledge-gap presupposed in a given explanatory context. The explanatory context, then, provides us with a kind of *metric* for pinpointing explanatory autonomy. Specifically, an explanation's autonomy is a measure of how well it closes knowledge-gaps compared to other explanations that might be given in some explanatory context.

For example, consider the following why-question: "Why does recall decay occur faster with multisyllabic words?" (Baddeley 2012). What is this question actually asking (i.e. the what is the explanatory context)? We begin by noting some general background assumptions made by Baddeley and colleagues (e.g. that working memory *has* a functional architecture and that understanding this architecture commits one to a certain set of research constraints).[29] The topic is explicitly stated by the question:

P_k = 'Recall decays faster with multisyllabic words.'

The contrast-class might plausibly consist of the following (among other things):

X = {'Recall decays faster with *monosyllabic* words'}.

[29] One background assumption about research constraints might be that certain experimental paradigms (e.g. observing the effects of distractors on list recall) are more relevant to the phenomenon of interest than others (e.g. single-cell recordings during list recall).

And, in Baddeley's case, the relevance relation may be whatever *psychological* factors are potential causes of the explanandum phenomenon, for example:

> *R* = Rehearsal capacity of the phonological loop; susceptibility of phonological memory to interference effects, etc.

Sticking to van Fraassen's PTE framework, an answer (*) to Baddeley's why-question given this explanatory context might be something like:

> (*) = Recall decays faster with multisyllabic words (rather than *monosyllabic* words) because longer words take longer rehearse and the phonological loop has a 2-second rehearsal threshold.

Assume that P_k, *X*, and *R* comprise a plausible explanatory context for Baddeley's interests, and further, that (*) is at least a candidate answer in this context, in the sense that it supports an understanding of the phenomenon such that it closes a relevant knowledge-gap.

Now imagine that a neuroscientist of memory expresses the same utterance that motivated Baddeley's explanation: "Why does recall decay occur faster with multisyllabic words?" The question arises whether this string of words expresses the same question as Baddeley's, i.e., whether it is requesting the same sort of understanding. Suppose the neuroscientist wants to identify and understand a neural mechanism underlying recall decay. As background, this neuroscientist assumes certain facts about cytoarchitecture (e.g. divisions of Brodmann areas) and prior research linking certain brain regions to language and memory tasks. Given the formulation of her why-question, this neuroscientist's question shares the same topic as Baddeley's. And for ease, suppose she presuppose the same contrast-class, i.e.:

> P_k = 'Recall decays faster with multisyllabic words.'
>
> *X*: {'Recall decays faster with *monosyllabic* words'}.

However, in relation to this topic and contrast-class, the neuroscientist is asking a different question from Baddeley's because there's a crucial difference in the relevance relation. The relevance relation, for the neuroscientist, pertains to the neural implementation underlying the observed effect on recall, perhaps something like:

> *R*': The neural structure whose entities, activities, and organization exhibit the effect.

Given her explanatory context, the neuroscientist may offer the following kind of answer:

> (*)' = Recall decays faster with multisyllabic words (rather than *monosyllabic* words) because mechanism *M* in the left posterior parietal region (an area responsible for short-term verbal recall) has a low peak activation threshold of *n* spikes/ms., and multisyllabic words generally produce spike trains that exceed this threshold in M.

Assume for the sake of argument that the contents of P_k, *X*, and *R'* are plausible ways of specifying the neuroscientist's explanatory context and that (*)' is at least a candidate answer in this context.

Since each researcher's why-question presupposes a different relevance relation, their answers address different aspects of the same overall phenomenon. Accordingly, these answers would support different forms of understanding and therefore fill different knowledge-gaps. Baddeley sought to fill a gap between an abstractly defined information-processing unit (the phonological loop) and a behavioral regularity. Our imaginary neuroscientist sought to fill a gap between a concretely defined neurobiological system (whatever *M* is a placeholder for) and that same behavioral regularity. These explanations are autonomous to the extent that differences in their respective explanatory contexts generate answers that close distinct knowledge-gaps by supporting distinct kinds of understanding.

5.3 Pragmatic autonomy, direct constraints, and explanatory relevance

At this stage, P&C might point out that discoveries of direct property constraints from lower-level brain mechanisms are often seen as relevant for developing functional analyses. For example, as Repovš and Baddeley note, memory psychologists often constrain their functional models according to discoveries about brain mechanisms:

> Human working memory is a system implemented in the brain and therefore constrained by its properties. Carefully planned and executed, studies that include the brain dimension can contribute important tests and insights to the development of a functional description of any cognitive ability. (Repovš and Baddeley 2006: 17)

Based on this passage, it seems undeniable that knowing about brain mechanisms can play a unique role in advancing our understanding of a cognitive capacity. But when is a direct neural constraint *relevant* to the explanatory power of a functional analysis? And how does the pragmatic framework on offer account for this?

A PTE framework can account for how our explanatory interests relate to certain ontic structures and not others (cf. Sirtes 2010). This framework's distinctive advantage is recognizing the role of human interests in determining explanatory relevance. For example, one main pragmatic interest of cognitive neuroscience is to link cognitive function to neural structures. In doing so this discipline abandons neutrality about which physical materials implement cognitive functions. Cognitive neuroscience thus has a particular stake in discovering neural constraints. The above quote from Repovš and Baddeley can be understood as presupposing a commitment to a relevance relation between "studies that include the brain dimension" and "the development of a functional description of any cognitive ability." But it is crucial to note that these constraint relations are relevant insofar as acknowledging them advances our pragmatic interests of closing distinct knowledge-gaps.

What pragmatic considerations go into determining when an ontic constraint is relevant for bringing about a certain phenomenon? As already discussed, one consideration is the level of abstraction at which an explanation is given in a particular context.[30] Once committed to studying a

[30] Sirtes (2010), who advances his own 'pragmatic-ontic' approach to mechanistic explanation, suggests that *levels* of concepts used to describe the phenomenon are pragmatically determined. I follow his lead on this point.

system at a certain level of analysis—a purely pragmatic decision—theorists can then assess relevance relations according to ontic standards of explanatory relevance. One such ontic standard is *causal relevance*, i.e., which causal variables in a system produce the phenomenon of interest. As stated earlier, one standard of causal relevance is *mutual manipulability*: the power of one causal variable A to change the value of another variable B in virtue of a change to A. Another ontic standard is *constitutive relevance*, i.e., which *parts* of a mechanism are responsible for producing the phenomenon of interest (Craver 2007b).[31]

Our pragmatic interest in explaining a phenomenon at a certain level of analysis will determine *which* class of manipulations exhibited by a system's causal variables are ones we should pay attention to. For example, cognitive neuroscientists often use transcranial magnetic stimulation to temporarily disable specific brain regions to see what effects this has on cognitive function. At this level of inquiry, manipulating cognitive function as a result of manipulating a brain region is a relevant manipulation because it pinpoints a causal variable involved in some phenomenon of interest (i.e. a cognitive capacity). Similarly, commitment to a level of analysis will determine which constitutive relations are relevant for producing a phenomenon of interest. So in asking which neural regions *make up* the episodic buffer in working memory, certain levels of neural architecture will be more relevant than other levels.

In thinking about how pragmatic factors help to determine explanatory relevance relations, it may help to consider a case where ontic constraints serve no explanatory role given one's pragmatic interests. P&C's claims about mechanistic constraints on general purpose computation can serve as the backdrop. They argue that "the task analysis of a general purpose computer does place

[31] Causal relevance and constitutive relevance are different in kind. As Craver notes, most views of causation typically require causes and effects to be "distinct and overlapping"; but the behavior of parts of mechanism and the behavior of the mechanism as a whole overlap (Craver 2007b).

direct constraints on its mechanistic explanation and vice versa" (p. 294). For them, this task analysis "describes precisely the special purpose circuit that is hardwired to generate the relevant computation" (p. 295).[32] P&C seem to overlook the possibility that a functional analysis may indirectly *describe* or *reference* physical hardware without that hardware being explanatorily relevant to some inquirer's why-question.

For example, a computer software technician looking to repair corrupted files on a RAMDisk software program may be *describing* a particular mechanism by asking about the corrupted files. Still, the factors most relevant to the purpose at hand—*repairing the RAMDisk*—might be far removed from that hardware. In search of a solution, she might map the RAMDisk technology on a flowchart to highlight potential sources of the problem. If, relative to her explanatory aims, the technician's map intelligibly connects some functional unit of RAMDisk to the target problem, this is what is explanatorily relevant for an explanation.

5.4 Is the pragmatic theory viable?

For all that I have said, it is open to anti-Autonomists to object that PTE is a failed theory. In particular, opponents of autonomy consistently emphasize that scientific explanations are not wedded to our pragmatic interests or to our psychology in any way. For instance, Craver (2014) suggests that pragmatic considerations are "downstream from discussions of what counts as an explanation for something else" (p.2, ms.). So why, the objection continues, think that a plausible conception of autonomy can be grounded in PTE?

A full defense of PTE is beyond the scope of this thesis. However, it is worth pointing out that the theory still has its defenders (e.g. Faye 2007; Sirtes 2010; Tschaepe 2009). Importantly, these

[32] For more explicit defenses of this view see Piccinini (2007, 2015).

defenders point out that, by bringing pragmatic considerations about explanation to the fore, researchers can better identify the problems they aim to study and therefore generate explanations that are suitable for solving those problems (again, think of the software technician troubleshooting the source of a corrupted file). Tschaepe makes just this point:

> Generally, pragmatic considerations are *those specific factors within an explanatory situation that determine what causes and effects are problematic and desirable, while being constrained by the exigence (or potential exigence) and its corresponding phenomenon. Pragmatic considerations entail the recognition that such considerations will lead to explanations that function as part of a solution or prevention.* These factors determine the premium on solving a problem, as well as the type of solution for solving that problem. (Tschaepe 2009: 37, original emphasis)[33]

Put this way, it is hard to see why paying close attention to the pragmatics of explanation would *not* be a central priority for anyone interested in seeking out explanations. The pragmatics of explanation help researchers to pinpoint, among other things, which sorts of factors (causal, constitutive, or otherwise) would be relevant for subjecting competing explanations to scrutiny: the winning explanation is the one that best accounts for all of the relevant difference-making features of the phenomenon of interest, relative to an explanatory context (cf. Strevens 2013, forthcoming).

My goal in proposing a pragmatic framework to account for autonomy was merely to specify what sorts of factors endow explanations with autonomy worth wanting. In fact, I think that PTE itself is dispensable. So long as pragmatic factors such as backgrounds, topics, contrast-classes, and relevance relations are given paramount consideration in the explanatory process, the Autonomist can stop short of a full endorsement of PTE.

But what about Craver's dismissal of pragmatic considerations as being "downstream from discussions of what counts as an explanation for something else"? Here again the ambiguity of

[33] 'Exigence' in this context simply refers to a problem in need of explanation, relative to a particular context.

'explanation' crops up. We can accept that, under the ontic mode, 'explanation' is devoid of pragmatic considerations: that the gating of ion channels explains calcium influx into a cell membrane is far removed from our own concerns. However, as I have argued above, science often attempts to bridge explanations *qua* causes (or constraints) with explanations *qua* models (representations, diagrams, etc.). And it is in this attempt to model explanations *qua* causes (or constraints) where the pragmatics of explanation become indispensable. Baddeley's model of working memory presumably attempts to capture the actual causal structure of that system on an abstract functional level (Weiskopf forthcoming). In doing so, however, the model is poised to close parochial knowledge-gaps.

Summing up, explanatory autonomy, in the sense that best reflects scientific practice, is an epistemic notion. That is, autonomy relates to the different contributions to knowledge and understanding that explanatory models make. By appealing to a pragmatic framework, I have suggested that our pragmatic interests highlight our particular epistemic deficits, and that the contents of a why-question indicate which sort of knowledge-gaps we wish to close. The examples of the syllable-length effect and the RAMDisk software aimed to highlight the fact that we can often ignore information about direct constraints without de-autonomizing the resulting explanation.

6 AUTONOMY WITHOUT DISTINCTNESS

So far I have taken issue with how P&C have characterized autonomy. But in doing so I have downplayed the role that their rejection of distinctness plays in undermining autonomy. I now want to argue that, even if the falsity of distinctness entails the falsity of P&C's no-direct-constraints autonomy, this need not undermine the version of autonomy I have defended.

6.1 The role of distinctness in rejecting and defending autonomy

Recall that P&C's argument against autonomy assumes of a strong relationship between distinctness and autonomy. They write:

> Distinctness is a necessary condition for autonomy: if functional analysis is *a kind of* mechanistic explanation, as we argue, then functional analysis cannot be *autonomous from* mechanistic explanation. (P&C 2011: 287-288, original emphasis)

Furthermore, P&C claim that other mechanists (e.g. Bechtel 2008a) "underestimate the role that distinctness plays in defenses of autonomy," and that, "[s]o long as distinctness remains in place, defenders of autonomy have room to resist the mechanists' objections" (p. 290).

P&C are right that defenders of autonomy have traditionally used distinctness as a crutch. However, I think P&C are mistaken in thinking that defenders of autonomy are *required* to endorse distinctness. One major exception is Bechtel (2007; 2008a; 2009), a mechanist about psychological explanation who endorses autonomy. I will briefly discuss Bechtel's view and then argue that my pragmatic framework helps explain why he is right.

Bechtel's approach to mechanizing psychological explanation, while also maintaining its explanatory autonomy, involves two methodological perspectives: "looking around" and "looking up" (Bechtel 2008a; 2009). By "looking around" a mechanism, researchers identify different ways in which components of a mechanism interact so as to exhibit phenomena of interest. This approach helps us to understand organizational principles that govern mechanisms and how these principles affect how a mechanism behaves. By "looking up" a mechanism, researchers aim to study how a mechanism behaves *in its natural environment* (e.g. how it responds to outside stimuli; how constraints in the environment impact its functioning). This approach helps us to understand how the "behavior of mechanisms is highly dependent on conditions in their environments" (2009: 559).

According to Bechtel, these two mechanistic perspectives highlight important sources of autonomous psychological explanations. First, regarding "looking around," Bechtel writes:

> [O]rganization enables systems of components to exhibit behavior different in character than that exhibited by the components. Such organized systems become the focus of their inquiry that is autonomous from the inquiry into the behavior of the components and focuses on how these systems engage the world in their own way. (Bechtel 2007: 191)

Autonomy, Bechtel argues, is rooted in part in how component organizations of a mechanism uniquely exhibit some phenomenon of interest. Different levels of component organizations uniquely exhibit certain phenomena of interest. And our epistemic divisions of labor attempt to fill knowledge-gaps created by a need to understand the phenomena that various levels of organization exhibit.

Furthermore, interest in a certain level of component organization narrows the scope of an explanatory context in significant ways. Most notably, different levels of organization will likely differ in relevance relation. For example, at one level, a mechanism might be *computationally* organized (e.g. partitions of working memory into abstract algorithms for encoding, storage, and retrieval). At another level, a mechanism might be *cytoarchitecturally* organized, intended to highlight the different cellular bases of neural networks. While elements of both of these levels may be constitutive of some overall mechanism (e.g. working memory), they surely regard different causes and constitutions as explanatorily relevant for their respective target phenomena.

The same point applies to the autonomous role that mechanisms play in relation to their *environments*. Given a set of interests about how some environment influences the behavior of a mechanism, theorists may state their why-questions in relation to unique relevance relations (e.g. causes in the natural environment) and contrast-classes (e.g. other environments or environmental constraints).

Bechtel sums up his construal of autonomy in a way consistent with my own. Specifically, he writes that "Independence [i.e. autonomy] stems from the fact that inquiry at each level provides information additional to that which can be secured at other levels, and generally does so using different tools of inquiry" (2008a: 157). The important point to note for present purposes is that such autonomy is consistent with psychological explanations being mechanistic in kind. While ontic constraints (e.g. component organizations of mechanisms) are acknowledged, these constraints are acknowledged to the extent that they exhibit phenomena relevant to our pragmatic aim of supporting certain kinds of understanding.

6.2 Against distinctness, for the multilevel mechanistic perspective

In the last subsection I argued that the rejection of distinctness is *compatible* with my pragmatic-epistemic conception of autonomy. In this subsection and the next I want to argue for something stronger: autonomous psychological explanation should *embrace* a multilevel mechanistic perspective on explanation. I begin by presenting a rationale for rejecting distinctness.

Some mechanists have attempted to reject distinctness by showing that psychology and neuroscience both traffic in *multilevel mechanistic explanation* (Bechtel 2008a; Craver 2007a). The multilevel perspective, recall, shows how each level of a common mechanism produces phenomena that contribute to the mechanism's behavior as a whole and also paints mechanisms as objects of interdisciplinary study (Craver 2007a).[34] With this framework, psychologists can employ methods, concepts, taxonomies, and models for explaining distinctively psychological phenomena, while also remaining steeped in the study of mechanisms (Bechtel 2008a, b).

[34] For this reason, the multilevel perspective is often seen as an alternative to traditional reductionism about scientific explanation (Bechtel and Abrahamsen 2005; Craver 2005, 2007; P&C 2011). As Craver puts it: "The reductive goal of globally relating two fields through derivation of laws should be replaced by a mosaic image of multiple fields making punctuate contributions to an abstract sketch of a mechanism" (2007: 246-247).

In my view, it makes sense to think that psychology and neuroscience traffic in multilevel mechanistic explanation and that, because of this, distinctness is false. Most importantly, both disciplines seem to posit structured entities and activities that are grounded in concrete mechanisms. By this I mean that entities and activities discussed by these disciplines are realized by physical systems that have mechanistic properties.

For example, researchers have identified neural structures thought to underlie Baddeley's phonological loop, which suggest a mechanism of action for the loop. First, two main functional components of the loop—phonetic storage and rehearsal—are thought to be grounded in distinct neuroanatomical regions within Broca's area (Awh et al. 1996; Paulesu et al. 1993). Second, storage and rehearsal are constitutive functions of verbal working memory, and since these functions are anatomically distinct, causal interactions between these two regions are necessary for verbal working memory. Third, if the frontal regions mediating rehearsal are spaced too far apart from parietal regions mediating storage, or if the projections between these two regions exhibit exceedingly long rates of information transmission or transcoding, verbal working memory as a whole would be impaired or absent.

My rejection of distinctness is open to the objection that psychological models are not mechanistic *in kind*. As Weiskopf (forthcoming) has argued, functional components of Baddeley's model "lack the characteristic properties of mechanistic entities" (p. 18).[35] Weiskopf's particular focus is on the apparent neutrality that this model has with respect to spatial properties.

However, it is not always the case that mechanistic cognitive models require specifications of *all* properties characteristic of mechanisms, as they are typically characterized in the literature. Even if the phonological loop lacks spatial properties, it still has other mechanistic properties, such

[35] This is true of Baddeley's model because no element of his analysis of working memory stands for a spatial property (e.g. shape) or relation (e.g. distance).

as having functional components that are plausibly based in mechanistic components (recall Awh et al.'s findings) and having causal variables that plausibly correspond to causal variables in neural mechanisms. Models seem to vary with respect to how many paradigmatic mechanistic properties they feature. Some feature all; some feature one or two. So why think that the absence of one feature renders the model non-mechanistic? The designation of a model as mechanistic may more plausibly be seen as a matter of degree.

Weiskopf could respond that I am talking about mechanism sketches, since, unlike complete mechanistic explanations (which specify *all* the relevant mechanistic details underlying some target phenomenon), sketches ignore many mechanistic details. I certainly do not want to maintain this. Indeed, I agree with Weiskopf that psychology has explanatory resources to give *complete* explanations.[36] In particular, I agree that cognitive models have resources to fill in missing details within an explanatory framework proprietary to psychology.

At this juncture, I think the pragmatic framework on offer helps us to see why the completeness of an explanation is relative to an explanatory context. For example, what might be a mechanism sketch for a neuroscientist might be complete for a cognitive psychologist. Explanatory context will cue us in on when an answer to a why-question has closed some relevant knowledge-gap in that context. Given these pragmatic considerations, judging a model to be a mechanism sketch is perhaps best seen as a byproduct of "looking around" to explanations from other disciplines for explanatory insights. In this way, a neural model of long-term potentiation might be seen as a sketch relative to the interests of a biochemist investigating biochemical mechanisms of memory.

[36] Sketches by definition are incomplete.

6.3 An example of multilevel mechanistic explanation

So far I have given a rationale for rejecting distinctness, suggesting that elements of cognitive models have plausible underlying neural mechanisms, even if these mechanisms vary with respect to how many paradigm mechanistic properties they exhibit.

I now want to give an example of multilevel mechanistic psychological explanation. The purpose of doing so is to reject distinctness by presenting a counterexample. With P&C, I think that the shape of a psychological explanation answers in part to the shape of neural mechanisms. However, I find their rejection of distinctness too demanding: we do not need to find strict mappings/correspondences to show that a psychological explanation is mechanistic. This will contribute to the main thesis of this paper by forcing proponents of the autonomy to secure autonomous psychological explanation on some basis other than distinctness.

Consider two models of edge detection—one computational-cognitive, the other neural. The first model is the Marr-Hildreth (1980) algorithm for edge detection in early vision. Though components of this model are *mathematically* characterized, it is still a functional analysis, since mathematical functions simply describe what each component of the model *does*.[37] The second model is Wu et al.'s (2007) spiking neural network model of edge detection.[38] This is a mechanistic explanation because it explains its target by describing relevant aspects of a neural mechanism. My claim is that both models describe a single mechanism and share a commitment to explicating a certain level of that mechanism.

---

[37] Relatedly, Egan (forthcoming) writes that "a functional analysis of a complex system may involve function-theoretic [i.e. mathematical] characterization" (p. 2, fn. 1, ms.).

[38] Spiking neural network models developed out of biologically oversimplified artificial neural network models, whose main drawback was a failure to reflect the time- and sequence-dependent dynamics of biologically-grounded neural systems (Ghosh-Dastidar and Adeli 2009). Compared to artificial models, these "biologically inspired" spiking models much more accurately reflect the dynamics of the firing patterns and connections among real-life spiking neurons (Meftah et al. 2010).

The main explanandum of both models is how the visual system computes light information in order to detect the presence of edges in the environment. In Marr and Hildreth's model, mathematically characterized entities such as zero-crossings and Laplacian and Gaussian gradient operators do the explanatory work. In Wu et al.'s model, directionally sensitive neurons and photoreceptors play a prominent explanatory role. Both models must explain edge detection in light of the same complicating factor: light intensity varies radically throughout any given retinal image, and without some way of constraining or averaging out these multifarious intensities, computing each value would become intractable and inefficient.

Each model gives a solution to this problem. Marr and Hildreth proposed a light filtration mechanism governed by a mathematical operator—a Gaussian filter. This filter smooths the image at various scales. By doing so changes in light intensity throughout each scale of the image are minimized (see Figure 2 below).[39]

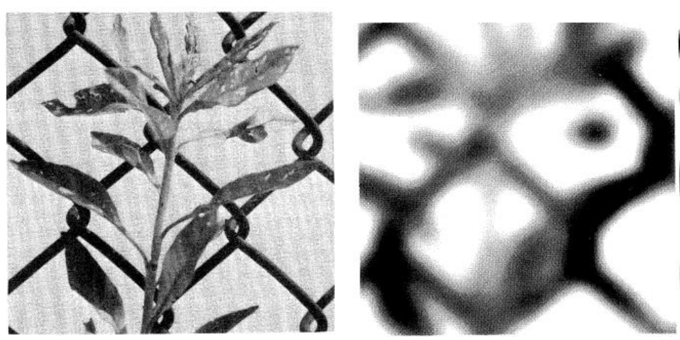

Figure 2 An example of Gaussian Filtration (Marr and Hildreth 1980)

After the Gaussian function is computed, and light intensity changes are minimized, a second function, a Laplacian, computes the Gaussian derivative (i.e. the light intensity changes after the Gaussian function has been performed), which is then assessed for *zero-crossings*.[40]

[39] This process is known as "convolving."

[40] Thus, the function known as the Laplacian of a Gaussian determines the likely presence of an edge by scouting for light intensity differentials that span the zero-crossings of a Gaussian-filtered image.

Wu et al.'s model (Figure 3) also solves the problem by positing an averaging mechanism.

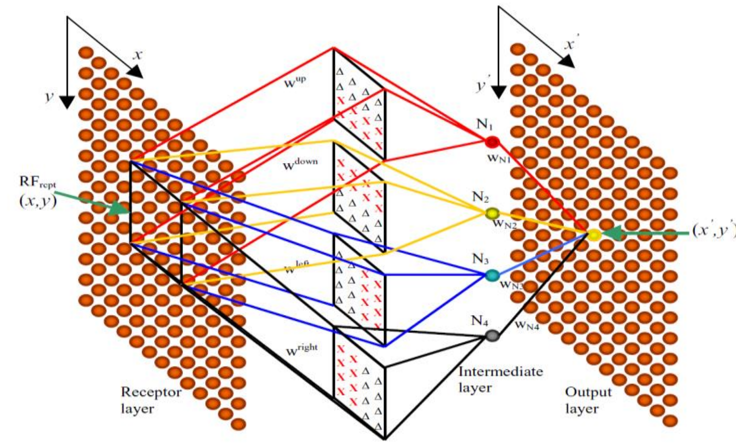

Figure 3 Wu et al.'s spiking neural network model of edge detection (Wu et al. 2007)

The input layer is a receptor site that receives photons. At the intermediate layer, each of the four tile-like cells correspond to a specific, directionally-specified (i.e. edge direction) type of neuron with a receptive field.[41] For example, when a photon impinges any given point on the receptor layer, that point will correspond to one of the four receptive fields. Finally, at the output layer, the inputs to each receptive field are integrated (as indicated by the single coordinate (*x*, *y*) at the output layer). Based on an averaging of all the inputs, the outer layer produces what Wu et al. call a "firing rate map," which "forms an edge graphic corresponding to the input image" (p. 27).[42]

---

[41] Red Xs and Δs represent excitatory and inhibitory connections, respectively.

[42] Other so-called "biologically inspired" models of edge detection have advanced this model by rendering the *shape* of their functional components similar to the actual shape of the physical entities that perform those functions. For instance, Kerr et al. (2011) develop a model of feature extraction whose photoreceptive layer mimics the shape of the hexagonal lattice structure of actual photoreceptors in the human fovea. As they put it: "Better spatial sampling efficiency is achieved by the hexagonal structure compared with a rectangular grid of similar pixel separation, leading to improved computational performance" (p. 382).

One can begin to see how these models might integrate, i.e., how they both describe different levels of the edge detection mechanism. The key is to see that loose mapping relations exist between the computational-mathematical functions of the Marr-Hildreth model and the spiking neurons in Wu et al.'s model. First, the Gaussian function in the Marr-Hildreth model corresponds roughly to the receptor layer in the SNN model. Second, the intermediate layer of Wu et al.'s model can be seen as computing the Laplacian function of the Gaussian derivative. Finally, the output layer of Wu et al.'s model can be seen as executing the search for a zero-crossing, based on averaged light information from previous processing or computation.

Here is my rationale for thinking this. If it is the receptor layer's job to process photons, and it is true that early vision creates a primal sketch in roughly the way Marr and Hildreth describe, then we should want some way of knowing how human photoreception *per se* does this. The function, then, is used not just as a mathematical description of a specific a cognitive capacity, but moreover, as a function to be carried out by a *neural mechanism*. The same is true of the intermediate and output layers: given our explanatory interests, there should be neural mechanisms that play the role of Laplacians and zero-crossing detectors.

This of course is a grossly oversimplified way of mapping a computational-cognitive model to a biological model. But my point here is not to iron out the details, but just to show how the models can begin to be described as multilevel mechanistic descriptions. So I am arguing, in the vein of P&C, that such model integration gives us reason to believe that distinctness, as construed by Autonomists, is false.

One could object, first, that all of this is merely to point out that all functions must be physically implemented. But it would not follow, the objection continues, that the functions are *mechanistically* implemented. For example, the neural bases of the Marr-Hildreth computations capacity may lack distinctively mechanistic properties such as start and finish conditions, regular operation, or spatial properties.[43] A second objection is that, as it stands, the Marr-Hildreth model lacks mechanistic properties.

In response to the first objection, much hangs on how we individuate mechanisms. There is reason to think that mechanisms are not natural kinds, but rather conventionally individuated systems that play important scientific roles (Craver 2009). For example, it might be that the neural computation that realizes the Gaussian function occurs regularly enough to warrant circumscribing some neural region and treating it *as* a mechanism. The relevant neural region might have predictable start and finish conditions depending on its sensitivity to various inputs and internal processes. The region might have parts that function as values of the arguments of the computation. And it might perform activities that constitute the type of mathematical operation that gets performed over those arguments and values (e.g. vector summation). So, my tentative conclusion is that if science has good reason to select a neural region as being uniquely responsible for a given cognitive capacity, and if that region respects mechanistic strictures, then it can fruitfully be treated *as* a mechanism.

As for the second objection—that the computational explanation of edge detection lacks mechanistic properties—it is plausible that the computations describe the conditions that a mechanism must satisfy in order to perform the relevant computations. For example, the Gaussian function is computable by many photosensitive physical systems, but all of these systems, different as they are, share properties that enable *that* function (rather than some other) to be computed in response to certain impinging stimuli. So, in Wu et al.'s model, the mapping of photoreceptors to arrays of

[43] Weiskopf makes this point with an analogy about the non-mechanistic nature of protein folding.

direction-sensitive neurons might be one instance of a more general spatial arrangement that enables *that* system to perform the Gaussian computation (if it does). For example, if the spatial arrangements (e.g. the surface area of each directional region) were manipulated, then the computation would change.[44]

7 CONCLUSION: PROSPECTS FOR INTEGRATION

In this thesis, I have attempted to argue, contra Gualtiero Piccinini and Carl Craver (2011), that autonomous psychological explanation fits into a mechanistic explanatory framework. I argued that mechanistic explanation does not threaten the kind of autonomy that defenders of the 'Received View' value. I suggested that autonomy is ineluctably tied to pragmatic and epistemic concerns about explanation. Specifically, autonomy depends on the unique epistemic contribution that an explanation makes to closing a relevant knowledge-gap in an explanatory context by supporting a certain kind of understanding of a phenomenon. Seen in this way, autonomy does not preclude the need for psychologists to have knowledge about neural structures when developing their explanations; it simply allows that, once developed, psychological explanations by themselves can uniquely support certain kinds of understanding.[45]

As stated above in §1, mechanists such as P&C view autonomy as inimical to the goal of integrating psychology and neuroscience. However, I argued that this worry should be directed only at *distinctness* (the claim that psychological explanations and mechanistic explanations differ in

[44] The link between photoreceptors and direction-sensitive regions in Wu et al.'s model is fixed via spatial weight matrices. So the manipulations I am thinking of are changes of these weights.
[45] This is similar to Weiskopf's claim (quoted above in §1) that autonomy is the ability for cognitive models by themselves to sufficiently explain their target phenomena. However, I add that these models have their autonomy in virtue of their unique epistemic contribution to closing relevant knowledge-gaps about target phenomena.

kind). In particular, I suggested that distinctness renders explanations in psychology and neuroscience *epistemically fragmented* by not supporting ways for researchers to integrate their explanations into a unified body of scientific knowledge. To the extent that one sees distinctness as necessary for autonomy, it is no surprise that autonomy would face the charge of epistemically fragmenting psychology and neuroscience as well.

P&C end their paper by summarizing their reasons for treating psychological explanations mechanistically:

> Given that psychological systems are in fact implemented in biological systems, and that such systems are more or less precisely replicated through reproduction, evolution, and development, there are frozen structural constraints on the mechanisms that do, as a matter of fact, implement behavioral and cognitive functions. Learning about components allows one to get the right functional decomposition by ruling out functional decompositions that are incompatible with the known structural details. (P&C 2011: 306)

This passage echoes an important point about *our* explanatory practices. While potentially many *kinds* of mechanisms can implement cognitive functions, the fact is that, in humans, such functions get performed by regularly operating biological mechanisms. It therefore makes sense to wed our interests in psychological phenomena to the mechanisms which, in our world, happen to perform them. It is in this sense that epistemic fragmentation between psychology and neuroscience can be avoided.

However, simply conceding to P&C that learning about neurobiological mechanisms is crucial for arriving at correct functional analyses need not force the Autonomist to retreat. The Autonomist can agree that functional analyses must be constrained by relevant knowledge about neural structure but still insist that the resulting functional analyses make unique epistemic contributions to our understanding of cognitive capacities.

P&C suggest that once functional analyses are supplemented with structural details, they lose their autonomy:

> The autonomist vision allowed experimental and theoretical psychologists to proceed with that task [of characterizing the cognitive phenomena for which neural explanations would be sought] without having to wait for neuroscience to catch up. Now the discipline has advanced to the point that these pursuits can meaningfully come together, and there are tremendous potential benefits from affecting such integration (P&C 2011: 308)

Doubtless, recent advances in cognitive neuroscience have brought the cognitive sciences closer to the study of the brain more than ever before—an interdisciplinary achievement worth applauding. Despite these advances, psychology and neuroscience can be meaningfully integrated without completely abandoning "the Autonomist vision."

The key is to pinpoint psychology's autonomy in how well it explains phenomena at a particular level within multilevel mechanisms. The pragmatics of explanation, I have argued, help to circumscribe what is explanatorily important about those levels, including their components, causal activities, and structural arrangements. Cognitive psychology has an established category of cognitive functions and operations that can be assimilated to mechanistic levels that are proprietary to psychology's taxonomy (Bechtel 2008b). Explanations couched at this level promote the pragmatic aims of psychology by uniquely illuminating phenomena that would otherwise go unaccounted for if neural-mechanistic details, once fully described, supplanted those categories and operations.

REFERENCE,S

Andersen, Holly (2014). “A Field Guide to Mechanisms: Part I.” *Philosophy Compass* 9 (4): 274-283.

Awh, Edward, et al. "Dissociation of Storage and Rehearsal in Verbal Working Memory: Evidence from positron emission tomography." *Psychological Science* (1996): 25-31.

--------(2012). “Working Memory: Theories, Models, and Controversies.” *Annual Review of Psychology* (63): 1-29.

Barrett, David (2014). “Functional Analysis and Mechanistic Explanation.” *Synthese* 191 (12): 2695-2714.

Bechtel, William & Abrahamsen, Adele (2005). “Explanation: A Mechanist Alternative.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36 (2): 421-441.

Bechtel, William & Richardson, Robert C. (1993). *Discovering Complexity Decomposition and Localization as Strategies in Scientific Research*. Princeton.

Bechtel, William (2007). “Reducing Psychology While Maintaining its Autonomy via Mechanistic Explanations.” In M. Schouten & H. L. De Joong (eds.), *The Matter of the Mind: Philosophical Essays on Psychology, Neuroscience and Reduction*. Blackwell Publishing.

--------(2008a). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. Routledge.

--------(2008b). “Mental Mechanisms in Cognitive Psychology: What are the Operations?” *Philosophy of Science* (75): 983-994.

--------(2009). “Looking Down, Around, and Up: Mechanistic Explanation in Psychology.” *Philosophical Psychology* 22 (5): 543-564.

Boone, Worth & Piccinini, Gualtiero (2016). “The Cognitive Neuroscience Revolution.” *Synthese* 193 (5): 1509-1534.

--------(Unpublished). “Mechanistic Abstraction.” Unpublished Manuscript.

Craver, Carl F. (2005). “Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience.” *Studies in History and Philosophy of Science Part C* 36 (2): 373-395.

--------(2007a). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press, Clarendon Press.

--------(2007b). “Constitutive Explanatory Relevance.” *Journal of Philosophical Research* 32: 3-20.

---------(2014). “The Ontic Account of Scientific Explanation.” In Marie I. Kaiser, Oliver R. Scholz, Daniel Plenge & Andreas Hüttemann (eds.), *Explanation in the Special Sciences: The Case of Biology and History*. Manuscript.

Cummins, Robert C. (1975). “Functional Analysis.” *Journal of Philosophy* (72) (November): 741-764.

--------(1983). *The Nature of Psychological Explanation*. MIT Press.

Egan, Frances (forthcoming). “Function-Theoretic Explanation and the Search for Neural Mechanisms,” forthcoming in Integrating Mind and Brain Science: Mechanistic Perspectives and Beyond, David M. Kaplan (ed.), Oxford University Press.

Faye, Jan (2007). “The Pragmatic-Rhetorical Theory of Explanation.” In Johannes Persson & Petri Ylikoski (eds.), *Rethinking Explanation*. Springer 43-68.

Feest, Uljana (2003). “Functional Analysis and the Autonomy of Psychology.” *Philosophy of Science* 70 (5): 937-948.

Fodor, Jerry A. (1968). *Psychological Explanation: An Introduction To The Philosophy Of Psychology*. Random House.

Ghosh-Dastidar, Samanwoy, and Hojjat Adeli (2009). "Spiking Neural Networks." *International journal of neural systems* 19.04: 295-308.

Hardcastle, Valerie G. (1996). *How to Build a Theory in Cognitive Science*. SUNY Press.

Illari, Phyllis (2013). “Mechanistic Explanation: Integrating the Ontic and Epistemic.” *Erkenntnis* 78 (2): 237-255.

Kaplan, David M. (2011).” Explanation and Description in Computational Neuroscience.” *Synthese* 183 (3): 339-373.

Kerr, D., Coleman, S., McGinnity, M., Wu, Q., & Clogenson, M. (2011). “Biologically Inspired Edge Detection.” In *Intelligent Systems Design and Applications (ISDA)*: 802-807.

Levy, Arnon (2013). “Three Kinds of New Mechanism.” *Biology and Philosophy* 28 (1): 99-114.

Marr, David, & Hildreth, Ellen. (1980). "Theory of Edge Detection." *Proceedings of the Royal Society of London B: Biological Sciences* 207.1167: 187-217.’

Meftah, Boudjelal, Olivier Lezoray, and Abdelkader Benyettou (2010). "Segmentation and Edge Detection Based on Spiking Neural Network Model." *Neural Processing Letters* 32.2: 131-146.

Paulesu, Eraldo, Christopher D. Frith, and Richard SJ Frackowiak. "The Neural Correlates of the Verbal Component of Working Memory." *Nature* 362.6418 (1993): 342-345.

Piccinini, Gualtiero (2006). "Computational Explanation in Neuroscience." *Synthese* 153 (3): 343-353.

--------(2007). "Computational Explanation and Mechanistic Explanation of Mind." In Francesco Ferrtti, Massimo Marraffa & Mario De Caro (eds.), *Synthese*. Springer: 343-353.

--------(2015). *Physical Computation: A Mechanistic Account*. Oxford University Press UK.

Piccinini, Gualtiero and Boone, Worth (forthcoming). "Mechanistic Abstraction."

Piccinini, Gualtiero & Craver, Carl (2011). "Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches." *Synthese* 183 (3): 283-311.

Repovš, Grega, & Alan Baddeley. "The Multi-Component Model of Working Memory: Explorations in Experimental Cognitive Psychology." *Neuroscience* 139.1 (2006): 5-21.

Roth, M. and Cummins, R. (Forthcoming). "Neuroscience, Psychology, and Functional Analysis." Manuscript.

Daniel Sirtes, "A Pragmatic-Ontic Account of Mechanistic Explanation", in: *PhilSci Archive 2010*, URL: http://philsci-archive.pitt.edu/archive/00005181/

Shapiro, Lawrence A. (2016). "Mechanism or Bust? Explanation in Psychology." *British Journal for the Philosophy of Science*: 1-23.

Strevens, Michael (2013). "No Understanding Without Explanation." *Studies in History and Philosophy of Science Part A* 44 (3): 510-515.

--------(Forthcoming). "Special-Science Autonomy and the Division of Labor." In Mark Couch & Jessica Pfeifer (eds.), *The Philosophy of Philip Kitcher.*

Tschaepe, Mark Dietrich (2009). "Pragmatics and Pragmatic Considerations in Explanation." *Contemporary Pragmatism* 6 (2): 25-44.

van Fraassen, Bas C. (1980). *The Scientific Image*. Oxford University Press.

Waskan, J. (2011). "Intelligibility and the CAPE: Combatting Anti-Psychologism about Explanation." In *Epistemology of Modeling & Simulation: Building Research Bridges Between the Philosophical and Modeling Communities*. Pittsburgh. Manuscript.

Weiskopf, Daniel A. (2011). “Models and Mechanisms in Psychological Explanation.” *Synthese* 183 (3): 313-338.

--------(Forthcoming). “The Explanatory Autonomy of Cognitive Models.” *Integrating Mind and Brain Science: Mechanistic Perspectives and Beyond*, David M. Kaplan (ed.), Oxford University Press.

Woodward, James (2003). *Making things happen*. New York: Oxford University Press.

--------“Scientific Explanation”, *The Stanford Encyclopedia of Philosophy* (Winter 2014 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=scientific-explanation

Wright, Cory D. (2012). “Mechanistic Explanation Without the Ontic Conception.” *European Journal of Philosophy of Science* 2 (3): 375-394.

Wu, QingXiang, et al. (2007). "Edge Detection Based on Spiking Neural Network Model." *Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*. Springer Berlin Heidelberg: 26-34.