

# Is Statistical Learning a Mechanism?

*Philosophical Psychology*

Riana J. Betzler

[rjb220@cam.ac.uk](mailto:rjb220@cam.ac.uk)

Department of History and Philosophy of Science

University of Cambridge

Free School Lane, Cambridge, CB2 3RH, UK

Abstract: Philosophers of science have offered several definitions of mechanism, most of which derive from biological or neuroscientific roots. In this paper, I consider whether these definitions apply equally well to cognitive science. I examine this question by looking at the case of statistical learning, which has been called a domain general learning mechanism in the cognitive scientific literature. I argue that statistical learning does not constitute a mechanism in the philosophical sense of the term. This conclusion points to significant limitations in the scope of the mechanist philosophy when it comes to accounting for explanation in cognitive science.

Acknowledgements: I would like to thank Tim Lewens, Anna Alexandrova, Chris Clarke, William Bechtel, and two anonymous reviewers for helpful comments on previous drafts of this paper. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC

Grant agreement no 284123.

Keywords: Mechanism, philosophy of psychology, statistical learning

Word count: 12,131

## 1. Introduction

William Bechtel and Adele Abrahamsen open one of their papers with the statement, “Cognitive science is, more than anything else, a pursuit of cognitive mechanisms” (Abrahamsen & Bechtel, 2006, p. 159). Bechtel and Abrahamsen are, among many others, proponents of what has been called the “new mechanist” philosophy of science. This “new mechanist” philosophy has been very successful in describing how explanation proceeds in sciences—biology in particular—that do not fit well under the deductive-nomological model standardly invoked by philosophers of science around the middle of the twentieth century. The statement that cognitive science pursues cognitive mechanisms suggests that at least some of the new mechanists think that their account of explanation can extend beyond biology, into cognitive psychology. In this paper, I will consider whether such an expansion could occur. What does it mean to give mechanistic explanations in cognitive psychology? Does the new mechanist philosophy apply as well to cognitive science as it does to biology?

I will examine these questions by considering the case of statistical learning, which represents a particularly good example of the kind of thing cognitive scientists investigate and illustrates what productive cognitive scientific research looks like. Furthermore, statistical learning is exactly the kind of thing that cognitive scientists themselves call a mechanism; they have explicitly referred to it as a domain general learning mechanism. I will evaluate whether statistical learning counts as a fully-fledged mechanism in the philosophical sense of the term by looking at prominent accounts of mechanism—with a special focus on those proposed by William Bechtel and Carl Craver. I argue that statistical learning does not currently count as a mechanism according to these views and hence, that there is a disconnect between mechanism in

cognitive science and mechanism in philosophy. I then consider the consequences of this disconnect for both cognitive psychology and mechanist philosophy.

Ultimately, I contend that part of the issue is terminological—cognitive scientists use the term ‘mechanism’ in a much looser sense than mechanist philosophers. But this is not the whole story. More importantly, the conclusion that statistical learning does not count as a mechanism in the philosophical sense of the term raises important questions as to the scope of the new mechanist philosophy. The example of statistical learning teaches us something significant about progress in cognitive science that many philosophers of mechanism have not accounted for. Cognitive scientists often do undertake a process that looks much like the mechanistic decomposition described by philosophers when they consider how a complex system, such as language or visual perception, works. In this sense, they seek mechanisms. However, they often arrive at phenomena, like statistical learning, that do not fit the criteria for mechanism set out by philosophers. They then give abstract characterisations of these phenomena. Although some philosophies of mechanism have begun to appreciate the role of abstraction in cognitive scientific explanation, others remain overly dismissive of its value; Craver and Bechtel diverge on this point. I will argue that the fact that these characterisations remain abstract does not mean that the process of explanation stops. Cognitive scientists continue to do much interesting and epistemically explanatory work in trying to understand how these abstractions operate. By dismissing these abstract characterisations as merely descriptive and not explanatory, some of the new mechanist philosophers miss out important aspects of cognitive scientific explanation.

In the first section of the paper, I will summarise philosophical accounts of mechanism in general. In doing so, I focus primarily on similarities among “new mechanist” views, although I also note where major divisions occur. In the second section, I will describe research to date on

the statistical learning “mechanism” in cognitive science. In the third section, I will evaluate the status of statistical learning—whether statistical learning counts as a mechanism according to the definitions provided by philosophers—Craver and Bechtel in particular. In the fourth section, I will consider the consequences of my findings for both cognitive science and the mechanist philosophy. I will conclude with a call for a broader account of mechanism. This is not to say that the new mechanists are wrong in highlighting the importance of mechanism; rather, they have sometimes been overenthusiastic about it to the detriment of understanding other important parts of the story of how cognitive scientists explain.

## 2. Philosophical Accounts of Mechanism

In philosophy of science, mechanism is a well-recognised concept, around which a vast literature has emerged. This approach has been especially productive in discussions of biology and neuroscience. More recently, there has been some interest in the role of mechanistic explanation in psychology, but there remains a great deal of work to be done in this particular area.

In this section, I will briefly review some recent work on mechanism. There are some important differences among the various “new mechanist” views, which Arnon Levy (2013) has articulated in an excellent taxonomy highlighting their commitments. One of the differences that has received a great deal of attention recently, and which will be relevant to this paper, concerns the various mechanists’ commitments regarding explanation (see e.g. Kaplan & Craver, 2011; Wright, 2012; Illari, 2013; Craver, 2014; Chirimuuta, 2014). Despite their differences, the new mechanist philosophies have major features in common.<sup>1</sup> As such, I will largely discuss the various accounts together in an effort to arrive at a generalized account of mechanism with

which to evaluate statistical learning. I will, however, draw on Levy's taxonomy and highlight differences among the new mechanists where relevant to the discussion.

### *2.1. Levy's Taxonomy: Three Central Mechanistic Theses*

Levy identifies three central theses that coexist within the mechanist literature but which are not always well-marked as separate commitments. He acknowledges that the extent to which various mechanist authors subscribe to these three theses differs. For the moment, I will simply reprise what the theses are. In the next section, I will go into more detail concerning how these theses play out in various authors' works.

(1) Causal Mechanism (CM): "[T]he view that causal relations, at least outside the domain of fundamental physical phenomena, exist in virtue of underlying mechanisms.<sup>2</sup> CM rivals other accounts of causation, such as regularity views. It is best seen as a contribution to metaphysics" (Levy, 2013, p. 100). Essentially, causal mechanism is a metaphysical view about the nature of causal relations. What it means to be causal is to be connected by a mechanism.

(2) Explanatory Mechanism (EM): "[A] thesis about explanatory relevance: it states that to explain a phenomenon, one must cite mechanistic information, i.e. specify underlying parts and their organisation. EM contrasts with other general accounts of explanation, such as the Deductive-Nomological model" (Levy, 2013, p. 100). Explanatory mechanism is, then, a view about what information is relevant to scientific explanation.

(3) Strategic Mechanism (SM): "[C]oncerns the cognitive-epistemic power of mechanistic modelling and related scientific methods. It asserts that certain phenomena are best handled mechanistically. Discussions of SM tend to construe 'mechanism' fairly narrowly, in machine-like terms" (Levy, 2013, p. 100). Strategic mechanism is a way of doing science, a way of approaching phenomena.

CM and EM are logically distinct theses; one may—and some do—subscribe to the view that causal relations exist in virtue of underlying mechanisms (CM) while denying that information about mechanisms is explanatorily relevant and vice versa. The relationship between EM and SM, however, is a bit more complicated and these views tend to run together in the mechanism literature. Failure to distinguish among these different theses generates confusion in contemporary debates about the applicability of the mechanism literature to various case studies in biology, neuroscience, and cognitive science. In the next section, I will draw heavily on Levy’s taxonomy to illustrate where the various mechanist philosophers diverge in their commitments, while also highlighting what they have in common.

## *2.2. Key Points in Mechanist Philosophy*

Most philosophers who are interested in mechanism focus on complex systems and the strategies for decomposing them. They have emphasised: (1) the importance of parts, entities, operations, activities, and their organisation; (2) that there are various levels of mechanistic explanation and that once a mechanistic decomposition “bottoms out” the process of explanation ends; (3) that the issue of mechanism is closely tied up with the issue of causality; and (4) that there is an important distinction to be made between mechanisms in nature and mechanistic explanations. I will briefly elaborate on these major features here.

### *2.2.1. Parts*

Most philosophical accounts of mechanism emphasise the importance of identifying the components of a mechanism, though they use different terms and have different conceptions of what counts as a component. Bechtel defines a mechanism as “a structure performing a function in virtue of its component parts, component operations, and their organization” (Bechtel & Abrahamsen, 2005, p. 423). Craver, along with his colleagues Machamer and Darden, defines

mechanisms as “entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions” (Machamer, Darden, & Craver, 2000, p. 3). Stuart Glennan, another mechanist, speaks of parts that interact according to causal laws: “A mechanism underlying a behavior is a complex system which produces that behavior by of [sic] the interaction of a number of parts according to direct causal laws” (Glennan, 1996, p. 52).

There are subtle differences between these accounts of components. One difference concerns the role that the parts are meant to play in the various mechanist conceptualisations (Levy, 2013). For some, like Glennan the parts interact “directly” thereby supporting causal dependencies. This notion of parts and their direct interaction contributes to Glennan’s ideas about the metaphysics of causation and constitutes a statement of Causal Mechanism (CM). For others, such as Bechtel and Craver, the parts simply play a role in scientific explanation (EM).<sup>3</sup> According to this view, scientific explanations should specify the components of a mechanism and their organization (Levy, 2013). Levy (2013) notes that it is often difficult to pick out exactly what counts as a part or a component: “Most advocates of EM have remained fairly coy about some of the key notions here. They give examples, but they do not provide explicit accounts of what a part is, how to construe organization, and which criteria pick out, among possible parts and organizational features, those that are explanatorily relevant” (p. 102). Craver is an exception to this in that he specifies that parts must have a stable set of properties and be detectable, that it should be possible to intervene on them, and that they should support manipulations of James Woodward’s description (Craver, 2007; Levy, 2013). Many proponents of explanatory mechanism (EM) offer a sophisticated description of how we go about finding the parts through processes of decomposition and localization, which following Levy’s taxonomy, can be seen as an expression of strategic mechanism (SM) (e.g. Bechtel & Richardson, 1993).

Another important difference in the way the mechanists construe parts concerns the degree of physicality and discreteness that the components are taken to have. Most mechanist views suggest that the parts must be physical, discrete, and spatially localizable. This is illustrated in the examples that they take from molecular and cellular biology, genetics and neurophysiology (Levy, 2013). Glennan (1996), however, contends that various kinds of entities can function as parts of a mechanism and that these parts need neither be spatially localisable nor describable in physical vocabulary. The inclusivity of Glennan's concept of mechanism means that it may be able to account for information processing mechanisms whose parts are software modules or data structures. However, one might also worry that Glennan's definition of parts is so broad as to be vacuous. Glennan tries to counter this worry by restricting what counts as a part to those things that "have a kind of robustness and reality apart from their place within the mechanism" but it remains unclear exactly what he means by this (1996, p. 53). Likewise, other mechanists such as Bechtel, have offered some descriptions of parts that downplay their need to be discrete and easily spatially localizable. For example, Bechtel and colleagues suggest that the parts may be distributed and non-discrete as in a connectionist network or dynamical model (Abrahamsen & Bechtel, 2006; Bechtel & Abrahamsen, 2010; Kaplan & Bechtel, 2011). This broader notion of parts or components, however, seems substantively different from the kinds of parts that Craver has in mind and it is often difficult, given these broader views, to see clearly what delineates a part from a non-part.

### *2.2.2. Levels and bottoming out*

Some of the mechanist philosophers, including Bechtel and Craver, describe mechanisms as hierarchical in structure such that larger mechanisms can be comprised of smaller ones. This raises the question of what the bottom level mechanisms might look like and whether our



explanatory powers extend to them. According to Craver, the lowest level mechanisms are “components that are accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group, or field” (Machamer, Darden, & Craver, 2000, p. 13). Explanation comes to an end when these bottom level mechanisms are found: “Bottoming out is relative: Different types of entities and activities are where a given field stops when constructing mechanisms. The explanation comes to an end, and description of lower-level mechanisms would be irrelevant to their interests” (Machamer, Darden, and Craver, 2000, p. 13).

On a related note, Glennan (1996) states that his description of mechanism applies equally well to all sciences except fundamental physics. In fundamental physics, there is no explanatory mechanism. Rather, the theories or laws postulated simply express brute facts about the world that cannot be mechanically decomposed (Glennan, 1996).

### 2.2.3. *Causality*

Another key feature of mechanist philosophy is the issue of causality, although the various accounts differ in the degree to which they view this issue as central. It is most central to the mechanistic accounts of Glennan and Craver, while it does not feature in Bechtel’s work.

The notion of causal laws is integral to Glennan’s definition of mechanism in that it allows for the interaction between parts. Glennan furthermore wishes to use his account of mechanism as the basis for a mechanical theory of causation as differentiated from a regularity theory of causation (Glennan, 1996, 2005, 2010). This mechanical theory of causation holds that the mechanisms are what form the “glue” holding entities together and that when we speak of one thing as *causing* another, we expect to find a mechanism connecting the two events (Glennan, 1996). Glennan gives the most direct expression of what Levy (2013) calls causal

mechanism (CM). It is the metaphysical thesis that what it means for one thing to cause another is for there to be a mechanism connecting the two (Levy, 2013).

Craver, in his later work, adopts James Woodward's counterfactual account of causation and incorporates it into his description of mechanistic explanation. In his book, *Explaining the Brain* (2007), he extends and modifies his earlier account of mechanism by fleshing out the meaning of the causal verb "productive" and uniting it with a view of causal relevance very close to that presented by Woodward (2002). Woodward's account of causation focuses on the process of intervention and how it allows for manipulation and control; simplifying greatly, it says that we think of causes as handles that allow of manipulation of events under idealised circumstances (Woodward, 2002). Furthermore, it states that the components of a mechanism should be modular in the sense that it should be possible to intervene on the behaviour of one component without necessarily altering the behaviour of the others. Craver calls his most recent view a causal-mechanical account of constitutive explanation. As this name suggests, his primary goal is to provide an account of scientific explanation (EM), not the metaphysics of causation. Constitutive explanation is the project of "understanding how the behavior of a whole is explained in terms of the behavior of its parts" (Craver, 2007, p. 160).

#### *2.2.4. Mechanisms in nature v. mechanistic explanations*

Though the ultimate goal is to have our mechanistic explanations correspond to the mechanisms found in nature, this is not always possible. In the long-term process of scientific research, one must go through many phases before this ideal goal can be attained. The various mechanist philosophers differ in their views on the status of these intermediate phases—whether they count as explanatory or not—and in their accounts of how one moves between these various phases. Levy's (2013) distinction between explanatory mechanism (EM) and strategic

mechanism (SM) becomes especially helpful here, as the two can be difficult to disentangle in many of the philosophers' views.

For Bechtel, the process of identifying a mechanism entails identifying its component parts and operations using methods of decomposition and localisation. According to this view, what it means to provide a successful explanation of a mechanism is to identify its component parts and operations, an expression of what Levy (2013) calls explanatory mechanism (EM). This endeavour may be difficult because smoothly functioning mechanisms conceal their operations (Bechtel & Richardson, 1993). Alongside EM, however Bechtel and colleagues' work also includes an expression of what Levy (2013) calls strategic mechanism (SM), or the view that approaching phenomena mechanistically is a good way of doing science. The SM aspects of Bechtel and colleagues' work can be seen in their sophisticated story about how we go about finding parts and operations through processes of decomposition and localization (e.g. Bechtel & Richardson, 1993). Furthermore, they emphasise the importance of representational devices that help researchers to visualise how the mechanism works (Sheredos, Burnston, Abrahamsen, & Bechtel, 2013; Bechtel & Abrahamsen, 2012). As Levy (2013) rightly points out, explanatory mechanism (EM) and strategic mechanism (SM) often "shade" together in Bechtel's work, perhaps because Bechtel, importantly, embraces the epistemic conception of explanation, where explanation is largely a cognitive, communicative activity aimed at generating understanding: "Providing explanations, including mechanistic explanations, is essentially a cognitive activity" (Bechtel & Abrahamsen, 2005, p. 425). This contrasts with the commitments of other mechanists—Craver in particular—who embrace an "ontic" conception of explanation. There has been much recent debate concerning the relative merits of the two conceptions of explanation, which I will leave aside for the moment (but see e.g. Kaplan & Craver, 2011;

Wright, 2012; Batterman & Rice, 2014; Bechtel & Abrahamsen, 2005; Chirimuuta, 2014; Illari, 2013).

In Craver's work, EM and SM remain much more separate. Craver's (2007) book, *Explaining the Brain*, focuses primarily on explanatory mechanism—the kinds of facts that are relevant to the explanation, including causal and constitutive relations, parts, and operations—but in some of his earlier work he does describe strategic mechanism—the process that scientists actually engage in when giving mechanistic explanations—as a separate enterprise. Craver (2006) describes the process of explaining a mechanism as involving stepping between *mechanism sketches*, *mechanism schemata*, and fully-fledged mechanisms. When a mechanistic explanation is incomplete, it is called a “mechanism sketch” and provides an account of “how possibly” a phenomenon comes about. Mechanism sketches may contain filler terms for parts of the mechanism that are not well understood. These filler terms may help in the scientific enterprise by temporarily standing in for a concept while other aspects of the model are elaborated but they can prove dangerous if investigators forget their transient status. A mechanism schema is a “truncated abstract description of a mechanism that can be filled with descriptions of known component parts and activities” (Machamer, Darden, & Craver, 2000, p. 15). They are descriptions that lack detail. According to Machamer, Darden, & Craver (2000), they play roles that are often attributed to theories in that they are “discovered, evaluated, and revised in cycles as science proceeds” (p. 16-17). Though Machamer, Darden, & Craver's (2000) description of schemata provides a role for abstraction, it does not view abstraction as genuinely explanatory.

Importantly, Craver endorses the “ontic” conception of explanation according to which explanations are objective and exist in the world to be discovered. This approach downplays the

epistemic or communicative role of explanations: “Ontic explanations are not texts; they are full-bodied things. They are not true or false. They are not more or less abstract. They are not more or less complete” (Craver, 2014, p. 40). On his view, an explanation reveals the mechanisms underlying the phenomena. In his work with David Kaplan, Craver expresses the importance of a “model to mechanism mapping (3M)” constraint, which states:

**(3M)** In successful explanatory models in cognitive and systems neuroscience (*a*) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (*b*) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism (Kaplan & Craver, 2011, p. 611).

Models of mechanisms, according to this view, are only truly explanatory when they map on to mechanisms in the world.<sup>4</sup> The 3M requirement has also been interpreted as implicitly, if not explicitly, endorsing a “more details the better (MDB)” assumption according to which the more accurate and detailed the model of a phenomenon is, the better it explains that phenomenon (Chirimuuta, 2014). This “more details the better” assumption has been critiqued extensively in recent work.

When it comes to their epistemic function, neither Bechtel nor Craver considers the intermediate steps in mechanistic explanation to be especially problematic. Bechtel and colleagues embrace mechanistic speculation as an essential component in the process of scientific discovery even if speculative mechanisms do not quite match up to mechanistic phenomena found in nature. Craver and colleagues, likewise, write that the mechanistic explanation helps to render the phenomenon intelligible whether it is correct or not:

The understanding provided by a mechanistic explanation may be correct or incorrect. Either way, the explanation renders a phenomenon intelligible . . . .Intelligibility arises not from an explanation's correctness but rather from an elucidative relation between the explanans (the set-up conditions and intermediate entities and activities) and the explanandum (the termination condition or the phenomenon to be explained) (Machamer, Darden, & Craver, 2000, p. 21).

However, the difference between the authors concerns whether the intermediate stages of mechanism construction can be counted as explanatory or not. For Craver, they most definitely cannot, as they do not (yet) adhere to the 3M requirement. Bechtel's more epistemic conception of explanation, on the other hand, makes room for the idea that even incomplete models of mechanisms may in some sense be explanatory.

Perhaps because of his strict adherence to the ontic conception of explanation, there is often a dismissiveness present within Craver's work about the value of other processes that we might be tempted to say go into successful explanatory work more generally. This dismissiveness is prevalent especially in his recent work (e.g. Craver, 2006). In a long description of the distinction between how-possibly and how-actually models, he gives the example of connectionist modeling and writes, of those models: "For those who merely want to predict the target system's performance, biologically implausible simulations will work just fine. But for those who build simulations in the search of explanations, mere simulation is not enough" (Craver, 2006, p. 361). He continues:

Philosophers of the special sciences . . . emphasize that explanations often proceed by functional analysis, reverse engineering, homuncular explanation, and decomposition. As a first-pass description of the nature of explanation of sciences such as cognitive

neuroscience, physiology, and molecular biology, this is a helpful descriptive framework. However, these accounts have yet to be developed to the point that they can distinguish how-possibly from how-actually functional analysis, reverse engineering, and homuncular explanation . . . . In the current view, constitutive explanations require descriptions of real mechanisms, not mere how-possibly posits (Craver, 2006, p. 362).

According to this statement, processes that may be involved in explanation, such as functional analysis, reverse engineering, and homuncular explanation provide *mere* how-possibly stories.

While Craver seems to suggest that these processes may eventually develop in ways that support how-actually models, they do not as of yet do so.

There is also, as Levy (2013b) points out, an incongruity between the tone and content of Craver and colleagues' discussion of the role of sketches and schemata. Levy (2013b) highlights this incongruity as it appears in Craver's (2008) discussion of the Hodgkin-Huxley model in particular. In that paper, Craver (2008) argues that the Hodgkin-Huxley model of the action potential is at most a sketch; it does not provide an explanation because it does not include molecular detail. Levy (2013b) writes:

The content and tone of Craver's discussion are not uniform. At some points he seems to evince a rather dismissive attitude, suggesting that the HH model has no explanatory content whatsoever. But his considered view, I think, is more moderate, namely that the model is a partial explanation, lacking in certain respects (p. 8-9).

As noted earlier, this dismissiveness may simply be down to Craver's endorsement of a very strict notion of explanation. I am not especially interested in getting involved in the debate over which conception of explanation is better here. However, as we will see later, the epistemic

conception endorsed by Bechtel affords a broader scope that may be better equipped to handle psychological and cognitive scientific mechanisms in particular.

### *2.3. Psychological Mechanisms*

The accounts of mechanism reviewed above were developed, for the most part, with biology and neuroscience in mind. I will now consider how they apply to psychology before evaluating whether statistical learning counts as a mechanism.

In his book, *Mental Mechanisms* (2008), Bechtel argues that cognitive scientists seek mechanisms. However, he also notes some important differences between biological and cognitive scientific mechanisms. One of the main differences is that explanations of biological mechanisms focus on material changes in physical substances whereas explanations of mental mechanisms focus on their ability to process information:

The performance of a mental activity also involves material changes, notably changes in sodium and potassium concentrations inside and outside neurons, but the characterization of them as mental activities does not focus on these material changes . . . . but rather on identifying more abstractly those functional parts and operations that are organized such that the mechanism can interact appropriately in its environment (Bechtel, 2008, p. 23).

Thus, according to Bechtel, when thinking about cognitive scientific mechanisms, we take an information-processing stance rather than a material stance.

Bechtel's recent work illustrates how the mechanistic philosophy might be applied to cognitive science and psychology. Although he is not committed to mechanistic explanations requiring neuroscientific detail, his account of mechanism nonetheless relies very heavily on neuroscience. Many of the examples in his book come from areas of cognitive science, including memory and vision, in which neuroscience has played an integral role in the process of



discovery. In these examples, he often points to parts that are discrete and spatially localizable. What happens when we are dealing with processes about which very little neuroscientific detail is known? Can mechanistic explanation still be productive in areas of cognitive science that depend on behavioural data? In cognitive science, identifying operations poses a major challenge because behavioural data provide very little information about the nature of the operations involved in transforming information (Bechtel, 2008). Furthermore, as Bechtel notes, decomposition in cognitive science often breaks phenomena into subtypes rather than operations. For example, in the case of memory, decomposition involved identifying various different types of memory, including declarative versus non-declarative memory and semantic versus episodic memory. Bechtel (2008) calls this a process of phenomenal decomposition rather than true mechanistic decomposition. Bechtel's solution to this involves neuroscience. He suggests that neuroscience data constrain cognitive scientific explanations:

The ability to identify brain areas with cognitive operations, though, provides a heuristic that is of great use . . . . This does not directly solve the problem of identifying operations but focuses attention on what might be operations invoked in these different psychological phenomena (2008b, p. 992).

Although Bechtel says that mechanistic explanation can occur without neuroscience, it is difficult to see exactly how.

Bechtel and Abrahamsen (2010) provide a fascinating account of how computational modeling contributes to dynamic mechanistic explanation in the cases of circadian rhythm research versus cognitive science. In it, they identify six contributions that computational modeling makes to the understanding of mechanisms. In the case of circadian rhythm research, computational models were constructed only after a great deal of biological detail was known

about the parts, operations, and organization of specific mechanisms. Cognitive scientific research is crucially different from circadian rhythm research, according to Bechtel and Abrahamsen (2010), because, “For the most part, empirical research in cognitive science has not revealed the representations or other component parts of cognitive mechanisms or their operations” (p. 329). This means that the computational model is often the only account of the mechanism in cognitive science and is less well-constrained than mechanistic models in other domains. Computational models in cognitive science, then, can make the same six contributions as computational models in circadian rhythm research. They can advance hypotheses and contribute much to the debate. But without an understanding of concrete parts and operations, conflicts among conflicting models cannot be resolved. Ultimately, Bechtel and Abrahamsen (2010) conclude that the use of computational modeling in cognitive science should emulate that in circadian rhythm research, despite the difficulty in doing so. This comparative discussion is helpful in that it shows how computational modeling, abstraction, and dynamic mechanistic explanation can be fruitfully and complementarily combined. This illustrates that Bechtel does embrace a broad view of how mechanistic explanation proceeds that is sensitive to the needs of cognitive scientific methods. But it nonetheless ends with a call to focus more on identifying the physical parts and operations of cognitive scientific mechanisms. I am in no way opposed to the search for parts of cognitive scientific mechanisms but, as will become clearer later in the paper, I do not think that productive work in cognitive science need proceed in precisely the same way as research in other areas of biology as exemplified by the structure of circadian rhythm research.

As noted earlier, Glennan thinks that his view applies equally well to all sciences, including psychology but excepting physics. His definition of parts is very inclusive and makes room for the possibility of non-physical parts and information processing mechanisms.

Furthermore, he provides an example of how his mechanistic philosophy might apply to the process of vowel normalisation in psychology. Vowel normalisation is the process by which speakers adjust to variations in the acoustical properties of different talkers' vowel sounds. In the case of vowel normalisation, the parts are difficult to localise. Glennan argues that it is nonetheless possible to construct a mechanical model even when one cannot directly identify and manipulate its parts (2005). Furthermore, he holds that it is possible to test competing models of the same phenomenon even without knowledge of the parts using methods of "indirect testing" (2005). But it remains unclear what kind of "indirect testing" Glennan has in mind and how it actually gets us closer to a mechanistic decomposition.

Craver and his colleagues' account remains closely centred on examples from biology and neurobiology and does not seriously consider psychological mechanisms as distinct from neuroscientific ones. In his work with Kaplan, he explicitly states that his view is meant to apply specifically to cognitive and systems neuroscience (2011). However, in a footnote, Craver and Kaplan (2011) say that they do not see why cognitive science should not be subject to the same mechanistic constraints as neuroscience. Hence, it seems that Craver and colleagues do, at least eventually, expect their account, including the 3M constraint, to apply to cognitive science.

Therefore, efforts to apply these mechanistic accounts to psychology have so far been limited. On the whole, they do not do much to specify how exactly mechanism can apply to higher-level cognitive scientific explanation. Bechtel and colleagues have done the most to extend mechanism to psychology and cognitive science and here I will attempt to extend these efforts further by looking at a specific test case from cognitive psychology: statistical learning. Statistical learning has been explicitly referred to as a mechanism by cognitive psychologists and is of particular interest because it has been investigated using primarily behavioural (rather than

neuroscientific) methods. In the next section, I will provide a fairly detailed description of the current research on the statistical learning “mechanism.”

### 3. What is statistical learning?

The term ‘statistical learning’ applies to several broadly related attempts to account for the *problem of inference* in human and machine learning. The problem of inference concerns how we make predictions, develop theories, construct models, and gain knowledge from a set of data (Bousquet, Boucheron, & Lugosi, 2004). Though statistical learning is of interest to researchers in several disciplines as a solution to various theoretical and practical problems, I will focus in this paper on just one instance of statistical learning—that thought to be involved in parsing continuous perceptual input into individuated units. I will give a fairly broad review of this body of research, both to illustrate the process by which investigations have progressed and to highlight the domain general nature of statistical learning. Focusing on the process by which statistical learning research has proceeded will help us to determine whether we are observing mechanistic decomposition at work. The domain general nature of statistical learning is also important, as will become clear later in this section, because it brought statistical learning itself, rather than the systems in which it is embedded, to the fore of research. The finding that statistical learning is domain general changed the direction of researchers’ attention and the nature of their explanatory project.

Historically, statistical learning research in cognitive science emerged out of investigations of the segmentation problem in language acquisition (Saffran, Newport, & Aslin, 1996). Spoken language is fairly continuous. People rarely pause between words and the pauses that they do make do not serve as reliable cues to word boundaries. Next time you hear a speaker of a foreign language, try to pick out individual words. It’s not easy and can lead to funny

segmentation errors: The French phrase *s'il vous plaît* consists of three words. Before I learned French, I thought it consisted of only one, especially because it translates to the single word 'please' in English. Children too make early segmentation errors: e.g. "I am hey/v/" from "behave" (Saffran, Newport, & Aslin, 1996). Interestingly, these sorts of segmentation errors happen much less frequently than one might expect and go away with linguistic experience. Solving the segmentation problem is one of the first steps we take in acquiring language but how we do so is a subject of much debate.

Saffran, Newport, and Aslin, in a series of papers published between 1996 and 1998, proposed that both adult and infant learners solve the segmentation problem by analysing the distributional cues to word boundaries present within linguistic input. They first demonstrated that human adults are able to segment an artificial speech stream in which the only cues to word boundaries are the transitional probabilities between syllables (Saffran, Newport, & Aslin, 1996). The transitional probability is a formal statistical measure indicating the frequency of co-occurrence of elements, computed as follows:

$$\text{Probability of X given Y} = \frac{\text{Frequency of pair XY}}{\text{Frequency of X}}$$

Because syllables that make up words are more likely to occur together than syllables that straddle word boundaries, word-internal transitional probabilities should be higher than word-external transitional probabilities. According to the statistical learning theory proposed by Saffran, Newport, and Aslin (1996), learners compute and compare these transitional probability statistics to discern the presence of word boundaries.

After their original demonstration that adult learners exploit transitional probability statistics to segment artificial speech streams—and are quite good at it too; the average mean score across subjects was 76%, a level significantly better than would be expected by chance—the authors extended these results to infant populations (Saffran, Newport, & Aslin, 1996). In their classic *Science* paper, they reported that infants as young as eight months of age can segment a continuous stream of artificial speech syllables into word-like units based on the statistical relationships between neighbouring sounds, even after only two minutes of exposure (Saffran, Aslin, & Newport, 1996). In a slightly later study, they established that the specific statistic being computed in this segmentation feat is the transitional probability (Aslin, Saffran, & Newport, 1998). It is important to note, of course, that these computations of conditional probability statistics must occur implicitly, below the level of conscious awareness (see Perruchet & Pacton, 2006 for a discussion). Otherwise, these results sound quite outlandish.

The language of mechanism pervades even these early papers on statistical learning. In their *Science* paper, Saffran and her colleagues explicitly refer to statistical learning as an experience-dependent mechanism: “Our results raise the intriguing possibility that infants possess experience-dependent mechanisms that may be powerful enough to support not only word segmentation but also the acquisition of other aspects of language” (Saffran, Aslin, & Newport, 1996, p. 1928). Though their interest, at this point, remained centred on language acquisition and perception, Saffran and colleagues do go on to suggest the possibility that the statistical learning mechanism may be domain general: “It remains unclear whether the statistical learning we observed is indicative of a mechanism specific to language acquisition or of a general learning mechanism applicable to a broad range of distributional analyses of environmental input” (Saffran, Aslin, & Newport, 1996, p. 1928). As statistical learning research

extended into other domains—vision in particular—the domain general nature of the statistical learning mechanism became increasingly clear.

The above-mentioned initial work stimulated further investigation into the limits of statistical learning for language acquisition and processing. One of the questions investigated concerned the nature of the representations generated by statistical learning: Does statistical learning simply generate strings of probabilistically related sounds or does the output of statistical learning have status relative to one's native language? Saffran (2001) demonstrated that infants treat the nonsense words extracted from an artificial grammar as word-like units in the context of their native language, English. Another series of papers investigated the extent to which statistical learning operates even in the face of complexities present in natural language, such as nonadjacent dependencies and words of varying length (Newport & Aslin, 2004; Johnson & Tyler, 2010); the constraints on statistical language learning (Peperkamp, Le Calvez, Nadal, & Dupoux, 2006; Saffran, 2002; see Saffran, 2003 for a review); and whether statistical learning operates even at the level of syntax (Peña, Bonatti, Nespor, & Mehler, 2002; Saffran & Wilson, 2003; Thompson & Newport, 2007; Bonatti, Peña, Nespor, & Mehler, 2005; Keidel, Jenison, Kluender, & Seidenberg, 2007). This work on the limits, constraints, and levels of statistical learning continues to have significant theoretical ramifications especially within the debate over the relative contributions of innate structures and learning mechanisms in language acquisition.

Further research demonstrated that statistical learning is not a uniquely human capability nor uniquely specified for language acquisition and processing. Hauser, Newport, and Aslin (2001) reported that cotton-top tamarins use statistical cues to segment streams of speech of the sort used in the original statistical learning studies. However, there are important differences between statistical learning in humans and statistical learning in tamarins; Newport and

colleagues demonstrated that tamarins, like humans, are sensitive to nonadjacent dependencies but over different elements (Newport & Aslin, 2004; Newport, Hauser, Spaepen, & Aslin, 2004). This work suggests that statistical learning may be constrained differently in humans than in non-human primates.

Outside the domain of language, statistical learning operates in non-linguistic audition, visual perception, action sequence segmentation, and causal perception (see Gopnik & Schulz, 2007, for detailed coverage of statistical learning in causal perception). The initial work on statistical learning in non-linguistic audition demonstrated that both human infants and adults succeed in using statistical cues to segment tone sequences (Saffran, Johnson, Aslin, & Newport, 1999). Later work on statistical learning of tone sequences centred on how it diverges in its specifics from statistical learning over linguistic input (Creel, Newport, & Aslin, 2004; Gebhart, Newport, & Aslin, 2009). Researchers investigating statistical learning of non-linguistic tones took the first step toward thinking of statistical learning as a domain general learning mechanism, which had important implications for its role in theories of language acquisition. However, perhaps because both linguistic and non-linguistic sounds occupy the auditory modality, more explicit recognition of the domain generality of statistical learning occurred only after investigation of its import in the visual domain.

Visual statistical learning research has been extremely fruitful. The body of empirical work on visual statistical learning is at least as large as that on statistical learning for language acquisition and theoretical discussion on its implications for larger theories of visual perception is just as lively. Kirkham, Slemmer, and Johnson (2002), upon demonstrating that 2-, 5-, and 8-month old infants can segment temporally presented sequences of shapes on the basis of transitional probability statistics, explicitly argued that statistical learning constitutes a domain



general learning *mechanism*; their title proclaims, “Visual statistical learning in infancy: evidence for a domain general learning mechanism.”

Further research in the visual domain demonstrated that learners extract both joint and conditional probability statistics from temporally presented visual streams, suggesting that statistical learning provides a viable mechanism for learning higher-order temporal structure, which may be implicated in event and action sequence perception in the real world (Fiser & Aslin, 2002; Baldwin, Andersson, Saffran, & Meyer, 2008). Another line of research investigates spatial visual statistical learning, which may be implicated in scene perception. Fiser and Aslin (2001) demonstrated that adult learners use both joint probability and conditional probability statistics to learn about spatial structures in visual scenes. This work has since been extended to show that infants too are capable of spatial visual statistical learning (Fiser & Aslin, 2002b). Spatial visual statistical learning remains effective even in the face of complex multi-element scenes and when categorizing real-world scenes (Fiser & Aslin, 2005; Brady & Oliva, 2008). Other work has examined the constraints on visual statistical learning (Fiser, Scholl, & Aslin, 2007); the role of attention in visual statistical learning (Turk-Browne, Junge, & Scholl, 2005; Toro, Sinnett, & Soto-Faraco, 2005); the elements over which it operates (Turk-Browne, Isola, Scholl, & Treat, 2008); and the interaction between spatial and temporal visual statistical learning (Turk-Browne & Scholl, 2009). Many of these questions and concerns are analogous to those investigated in the language domain.

Visual statistical learning researchers, like auditory statistical learning researchers, invoke the language of mechanism. They use many of the same methodologies, constructing their studies so that they can be directly compared to those conducted in the auditory domain, sometimes running visual and auditory analogues of the same experiments in parallel. These

aspects of visual statistical learning research, along with explicit statements about the domain generality of statistical learning (e.g. Kirkham, Slemmer, & Johnson, 2002), suggest that many researchers think of statistical learning as one central mechanism operating across many contexts. However, it is possible that there are multiple statistical learning mechanisms that all operate somewhat similarly. Accounting for this possibility, Brady and Oliva (2008), refer to statistical learning mechanisms in the plural: “The purpose of the present study was to examine the extent to which statistical-learning mechanisms can operate automatically at an abstract, conceptual level” (p. 679). Whether statistical learning consists of one or many mechanisms remains an open question, one that may in the future be resolved. Most work to date implies that researchers generally think of statistical learning as a single mechanism that applies across a broad array of contexts. The work that does refer to statistical learning mechanisms in the plural seems to construe the potential differences between mechanisms in terms of ability to compute different statistics over different regularities, having different constraints, or operating in different domains. These authors also sometimes tend to slip back and forth between referring to statistical learning in the plural and in the singular, and it is not clear how seriously researchers have considered the matter (e.g. Saffran, 2003).

In any case, the above review illustrates that the focus of statistical learning research has been on its phenomenological properties, constraints on when and how it operates, what kinds of representations it generates, and how it interacts with and is embedded in other learning systems. There has been very little effort to undertake a full decomposition of the mechanism in the new mechanist sense of the word.

Very recently, there has been some effort to uncover the neural correlates of statistical learning. Even this work in neuroscience, however, is not centrally concerned with decomposing

the statistical learning mechanism into parts or smaller mechanisms. It remains in the same vein as prior behavioural work. Turk-Browne and his colleagues used fMRI to investigate how statistical learning relates to other types of learning, how it affects perceptual processing, and how quickly it occurs (Turk-Browne, Scholl, Chun, & Johnson, 2009). They found sensitivity to statistically structured sequences in the right caudate of the striatum, which suggests that statistical learning may be related to other forms of associative learning. They also found activity in the medial-temporal lobe in the hippocampus, which suggests that statistical learning may be related to contextual learning, category learning, sequence learning, and relational binding. The authors write that the involvement of both the caudate and the hippocampus indicates that statistical learning may yield parallel abstract and specific representations; learning involving the hippocampus is abstract and may generalise easily to new contexts while learning involving the caudate is more specific and may only be expressed in situations that replicate the encoding context (Turk-Browne, Scholl, Chun, & Johnson, 2009). Finally, the fMRI data indicated that learning happens very quickly and implicitly. Neural responses to statistical structure were robust even when behavioural data were weak.

To summarise, statistical learning is currently thought of, by most researchers, as a domain general learning mechanism. Abstracting away from the empirical work, the statistical learning mechanism might operate as follows: (1) continuous streams of perceptual input (i.e. speech, visual scenes) containing distributed regularities are perceived by a modality (i.e. vision, audition); (2) the statistical learning “mechanism” processes the input by computing conditional probability statistics among various elements of the input (i.e. phonemes, features, shapes, syllables); (3) discrete units (i.e. words, objects) emerge from out of the continuous input. The statistical learning mechanism thus parses or segments the continuous input into discrete

elements. These elements might undergo further processing by the same kind of mechanism operating at a different level. Or they might be processed or used in other ways by different cognitive mechanisms. The mechanism is domain general in that it operates in essentially the same way over similar kinds of perceptual input across distinct domains though it may produce very different kinds of knowledge depending upon the specific regularities fed into it (Saffran & Thiessen, 2007).

What the statistical learning mechanism actually does remains rather obscure. We can say that it computes conditional probability statistics but questions remain as to how, where, and under what constraints. Currently, much of our knowledge of how statistical learning works rests on behavioural data. Investigations of its neural underpinnings have begun only recently. Because of the nature of the imaging methods used, these neuroscientific studies cannot give us much detail about the inner workings of statistical learning. Can statistical learning, despite the obscurity in the specifics of how it operates, still be called a mechanism?

#### 4. What is the status of statistical learning? Is it a mechanism?

In this section, I will consider whether statistical learning can fairly be called a mechanism based on the definitions provided in the philosophical accounts reviewed above. To recap, in general philosophical accounts of mechanism and mechanistic explanation emphasise the importance of identifying the components of a mechanism, though they use different terms and have different conceptions of what counts as a component. Both Bechtel and Glennan construe “parts” broadly so as to accommodate those that constitute information processing mechanisms. Craver’s account is deeply grounded in its biological and neuroscientific roots, meaning it may be more difficult to apply to cognitive phenomena. Furthermore, philosophers of mechanism emphasise the importance of recognising the distinction between mechanisms found

in nature and the process of mechanistic explanation. They also have different commitments concerning what counts as an explanation, which will be important to keep in mind throughout this section.

If statistical learning is to be called a mechanism under any of these accounts, we must be able to describe its components. Doing so, however, proves rather difficult. If there are parts in the statistical learning mechanism, they cannot, as of yet, be physically seen. Most of the research on statistical learning has been done using behavioural methodologies. Many of these methodologies are quite cunning and have allowed researchers to make real progress in probing psychological phenomena. But these behavioural methodologies do not offer information about isolated components, physical or not. Furthermore, though some researchers have begun to identify the neural correlates of statistical learning (e.g. Turk-Browne, Scholl, Chun, & Johnson, 2009), this project remains a long way away from being able to isolate specific entities. Identifying neural correlates is not sufficient for identifying the parts of a mechanism.<sup>5</sup> We do not yet know if the statistical learning mechanism as it exists in nature has components. In this sense, we cannot yet call it a mechanism on any of the new mechanist accounts described above. The closest we get, in the existing literature, to an account of mechanistic explanation that might include statistical learning mechanism(s) is Bechtel and Abrahamsen's (2010) discussion of dynamic mechanistic explanation and the role that abstract computational mechanisms play there. On Bechtel's view, it might very well be a *de novo* proposal of a mechanism that may be further given empirical grounding in the future.

But whether or not the statistical learning mechanism actually has components is not my concern here—this is an empirical question. I am more interested in whether researchers who investigate statistical learning are really trying to articulate parts—whether the articulation of

parts is an integral part of the research heuristic at work. This bears on whether they are undertaking a mechanistic explanation regardless of whether or not they actually find a fully-fledged mechanism in the philosophical sense of the term.

Researchers initially conceived of statistical learning as a component in a mechanism of language acquisition. They characterised it using the formal mathematical equation for calculating the transitional probability statistic. However, as further research demonstrated the domain generality of statistical learning, investigators began to speak of it as a mechanism in its own right. As Bechtel points out when he suggests that mechanisms are recursive, there is nothing problematic about this; statistical learning can function both as a mechanism in its own right and a constituent of a larger mechanism (Bechtel & Abrahamsen, 2005). But thinking of statistical learning as a mechanism in its own right would mean moving beyond the mathematical formalisation in our characterisation of it and identifying its parts.

In the case of statistical learning, researchers have been much interested in the role of other cognitive mechanisms, such as attention, in guiding its operation. This kind of research, as well as that on the limits and constraints of statistical learning, tells us a great deal about how statistical learning operates and gives us clues as to where in the brain it might be located without telling us anything about the parts of which it may be made up. Despite all of the research on statistical learning, it remains a black box. If we tried to draw a diagram of the statistical learning “mechanism,” it might look something like this:

[FIGURE 1 here]

But as is the case for many other psychological phenomena, this diagram doesn't tell us much. Though diagrammatic representations serve as very effective forms of communication in neuroscience and biology, we rarely find them in cognitive science (Bechtel & Abrahamsen,

2005). Eventually, the black box might be broken up into its components and the diagram might begin to look more like the one for circulation of blood through the heart. But the search for parts or components has not been the primary aim of research on statistical learning.

Thus, researchers have not, as of yet, identified differentiated parts, entities, components, or activities for the statistical learning mechanism. This means that it is also difficult, as of now, to determine whether the components of statistical learning (if there are any) are in fact governed by a generalisation that is invariant under interventions or causal laws, or are modular, as Glennan (1996), and Craver (2007) require. So statistical learning does not look as though it counts as a mechanism in the modern philosophical sense of the term. Researchers who investigate statistical learning have not undertaken a full mechanistic decomposition and as such, we do not see a mechanism—or at least not a fully-fledged mechanism. It looks like we may have a mechanism sketch on Craver's view, or a *de novo* proposal of a mechanism on Bechtel's view.

But what are the consequences of this conclusion? What does this mean for the philosophy of mechanism? What does this say about the use of the term mechanism in cognitive science? Many suggest that cognitive science is a pursuit of mechanisms (e.g. Kaplan & Craver, 2011; Bechtel & Abrahamsen, 2010; Abrahamsen & Bechtel, 2006). In what sense can we say that this is truly the case?

## 5. Consequences of Statistical Learning's Non-Mechanistic Status for Cognitive Science and Mechanist Philosophy

The conclusion that statistical learning is not a mechanism has several consequences for both cognitive psychology and mechanist philosophy. It could mean: (1) that the mechanist philosophy does not translate well from biology and neuroscience to cognitive psychology; (2)

that there is a terminological issue associated with the word ‘mechanism’ and that cognitive scientists and mechanist philosophers just aren’t using the term in the same way; or (3) that the statistical learning mechanism is not a mechanism at all but rather something else entirely. I believe that we are seeing a combination of these three things.

Part of the problem is that cognitive scientists and mechanist philosophers are not using the term ‘mechanism’ in the same way. The mechanist philosophers have developed very detailed definitions of mechanism with highly specific criteria as to what counts. These definitions usually are meant to describe complex systems consisting of multiple levels and parts. Cognitive scientists, on the other hand, seem to be using the term mechanism in a much looser sense. They use the term mechanism simply to refer to the means by which some form of behaviour is produced. Their definition of mechanism does not carry all of the trappings of the philosophical concept of mechanism. There is a lack of congruence in the terminology used by cognitive scientists and that used by philosophers. While this point is worth making, it is not especially profound.

More significantly, there is a real sense in which statistical learning researchers are not describing a mechanism. Statistical learning may turn out to be a fully-fledged mechanism in the future, if and when further work is done to articulate its parts, operations, and their organization. But for the moment, it provides an abstract characterization of what looks to be a domain-general phenomenon. At present, abstraction is central to the work of cognitive scientists in investigating how the statistical learning mechanism works and the search for parts and their localization takes a back seat to this abstraction. Abstraction plays an important role in various sciences, mathematics and computer science in particular. In computer science, abstraction is taken to allow us to appreciate the similarities between objects, situations, or processes (Turner, 2014). In



abstracting, computer scientists ignore concrete features and structural aspects of the phenomenon and in doing so arrive at a new concept. By calling statistical learning an abstract characterisation, I hope to point out that it does something similar to abstractions in computer science. It leads us to appreciate similarities between various different processes, such as linguistic and visual perception, and in doing so allows us to conceptualise phenomena in new ways, changing the kinds of questions we ask about those phenomena. Furthermore, statistical learning is often represented in the form of mathematical formalisation, highlighting its abstract nature. It is its abstract nature that allows it to do especially productive work in cognitive science.

Very recently, issues of abstraction have come to the fore of discussions of mechanism. In a response to Craver's (2008) paper, "Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential," Levy (2013b) argues for the importance of abstraction in biology. According to Craver (2008), the Hodgkin-Huxley equation is explanatorily deficient because it does not reveal enough detail about the underlying molecular specifics of the action potential. He argues that even if there are laws in biology and the Hodgkin-Huxley equation describes one, it is the biological facts and not the laws that do the explaining. Levy (2013b), contra Craver, provides a reading of the Hodgkin-Huxley case in which they deliberately abstract from molecular specifics to explain whole-cell behaviour. This kind of abstraction, Levy argues, is sometimes incredibly beneficial and explanatory, though it has been neglected by many proponents of mechanism: "Mechanism has a lot going for it but, at least as developed to date, it has tended to overemphasize the description of concrete parts and their spatiotemporal organization. Explanatory strategies that do not operate in this fashion have been overlooked, despite their importance in many parts of biology" (Levy, 2013b, p. 2). In a

further paper written with William Bechtel, Levy elaborates on the role of abstraction: “Tersely put, abstraction is the omission of detail. An abstract description includes only some of what could, in principle, be said about its subject matter. It leaves matters open, in certain respects” (Levy & Bechtel, 2013, p. 242). Abstraction, in Levy and Bechtel’s (2013) view, helps us to understand the organisation of the mechanism, but its role in biology has been under-appreciated. Chirimuuta (2014) has likewise raised important critiques of the mechanist philosophy as applied to computational neuroscience and emphasised the importance of abstraction there. In cognitive science, the role of abstraction is perhaps even more important. Although abstraction away from concrete detail may be deliberate in some of the biological sciences, in the cognitive sciences such abstraction is often necessary as we do not yet have the methods to probe such structural detail.

Abstraction may seem to be at odds with mechanistic decomposition but through the example of statistical learning, we see that the two explanatory procedures often coincide, and support one another, in the cognitive sciences. Interestingly, in the case of statistical learning, it was a process of decomposition—or what cognitive scientists viewed as a process of decomposition—that led to abstraction. Cognitive scientists viewed themselves as engaged in a search for mechanisms, as evidenced by their frequent use of the language of mechanism. By trying to decompose the system of language perception, they discovered statistical learning, a phenomenon whose abstractness led them to appreciate the similarity between various perceptual processes. Focusing on statistical learning as an interesting point of inquiry in its own right opened up new questions about these parallel perceptual processes.

Thus, both decomposition and abstraction are part and parcel of the process of explanation in cognitive science.<sup>6</sup> Much of the mechanism literature has been overly dismissive

of forms of explanation that do not capture the concrete details of parts, entities, and organisation, and thereby remains limited in the kinds of scientific endeavours to which it can apply. This is especially the case for Craver's view, on which explanation must be ontic and for which more details are better. If Craver hopes to extend his views into cognitive psychology, he will need to broaden the ambit of what he considers to count as explanatory to include higher-level explanations as can be found in cognitive science. Other mechanists, such as Bechtel, have made more room for cognitive science in their accounts of how explanation proceeds. For his account, I am simply contributing a minor tweak in the way in which abstraction and computational modeling are construed in relation to processes of mechanistic decomposition.

## 6. Conclusion

In this paper, I reviewed some contemporary accounts of mechanism that exist in the philosophical literature. These philosophical accounts of mechanism were developed with neuroscience and biology in mind but the proponents of these accounts hope that they may extend to cognitive science and psychology as well. I then described a phenomenon, statistical learning, that cognitive scientists refer to as a "mechanism." I tried to determine whether the cognitive scientific "mechanism" statistical learning fits the description given by philosophers. I found that it does not. This conclusion has several important consequences for cognitive science as well as for the mechanist philosophy. It suggests that there is a discrepancy between cognitive scientists and philosophers in the way that "mechanism" is being used. But more importantly, it indicates that there are instances in which the search for a mechanism does not actually uncover a mechanism but rather leads to an abstraction. Cognitive scientists do productive work as they articulate the role of these abstractions. The mechanist philosophy has been very fruitful in capturing some of the major drivers of research in biology, neuroscience, and cognitive science

alike and their views might be further expanded to capture the myriad ways in which cognitive scientists explain.

Notes:

---

<sup>1</sup> Note that this is consistent with how the “New Mechanists” have been viewed in the wider literature. Cory Wright, for example, highlights the similarity among the New Mechanists even when it comes to one of the most controversial issues—explanation: “Pending any such head count [regarding who counts as a New Mechanist], there is still a visibly high degree of unity and standardization among discussants of the proper conception of explanation in the literature on MC” (2012, p. 378).

<sup>2</sup> Levy (2013) is not committed to the idea that underlying need mean “smaller” and provides some discussion of Darden’s (2006) work on this subject.

<sup>3</sup> They differ on specifically what it means to explain, as I will discuss later, but parts are still thought to feature in explanation on all of the accounts.

<sup>4</sup> It is not clear the extent to which this constraint is meant to apply to cognitive science, an issue which I will discuss in the next section of this paper.

<sup>5</sup> It is worth noting that neuroscience includes various methodologies that afford a whole range of different levels of specificity. Non-invasive methods such as fMRI and ERP provide coarse-grained information about the brain and it is these methods that have been used in the case of statistical learning. Invasive methods such as single-cell recording allow for fine-grained information about the brain but for ethical reasons are not suitable for use in humans. It is these physiological methods that provide the kinds of information about parts, entities, activities, and components that the new mechanists seek. Thus, the scope of mechanism may be constrained even within neuroscience. Levy (2013) makes a similar point with regard to biology; he notes that while the mechanist philosophy has been successful within areas of proximal biology, it may be less so in evolutionary biology.

<sup>6</sup> I am using explanation in an epistemic sense here.

## References

- Abrahamsen, A. & Bechtel, W. (2006). Phenomena and mechanisms: putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R. Stainton (Ed.), *Contemporary Debates in Cognitive Science* (pp. 159-186). Oxford: Blackwell.
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9(4), 321-324.
- Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, 106(3), 1382-1407.
- Batterman, R. W. & Rice, C. C. (2014). Minimal Model Explanations. *Philosophy of Science*, 81(3), 349-376.
- Bechtel, W. (2008). *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*. New York & London: Taylor & Francis.
- Bechtel, W. (2008b). Mechanisms in cognitive psychology: What are the operations? *Philosophy of Science*, 75(5), 983-994.
- Bechtel, W. & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421-441.
- Bechtel, W. & Abrahamsen, A. (2010). Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science*, 41, 321-333.
- Bechtel, W. & Abrahamsen, A. (2012). Diagramming phenomena for mechanistic explanation. *Proceedings of the 34<sup>th</sup> Annual Conference of the Cognitive Science Society* (pp. 104-107). Austin, TX: Cognitive Science Society.
- Bechtel, W. & Richardson, R. C. (1993). *Discovering Complexity*. Princeton: Princeton

University Press.

- Bonatti, L. L., Peña, M., Nespors, M. & Mehler, J. (2005). Linguistic constraints on statistical computations. *Psychological Science*, 16(6), 451-459.
- Bousquet, O., Boucheron, S., & Lugosi, G. (2004). Introduction to statistical learning theory. In O. Bousquet, U. v. Luxburg, & G. Rätsch (Ed.), *Advanced Lectures in Machine Learning* (pp. 169-207). Berlin & London: Springer.
- Brady, T. F. & Oliva, A. (2008). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychological Science*, 19(7), 678-685.
- Chirimuuta, M. (2014). Minimal models and canonical neural computations: the distinctness of computational explanation in neuroscience. *Synthese*, 191, 127-153.
- Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: a connectionist model. *Language and Cognitive Processes*, 13(2/3), 221-268.
- Craver, C. (2006). When mechanistic models explain. *Synthese*, 153(3), 355-376.
- Craver, C. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.
- Craver, C. (2008). Physical law and mechanistic explanation in the Hodgkin and Huxley model of the action potential. *Philosophy of Science*, 75(5), 1022-1033.
- Craver, C. (2014). The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, A. Hüttemann (Eds.), *Explanation in the Special Sciences: The case of biology and history* (pp. 27-52). Berlin: Springer.

- Creel, S. C., Newport, E. L. & Aslin, R. N. (2004). Distant melodies: statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(5), 1119-1130.
- Darden, L. (2006). *Reasoning in Biological Discoveries*. New York: Cambridge University Press.
- Fiser, J. & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science*, 12(6), 499-504.
- Fiser, J. & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 458-467.
- Fiser, J. & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *PNAS*, 99(24), 15822-15826.
- Fiser, J. & Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology: General*, 134(4), 521-537.
- Fiser, J., Scholl, B. J., & Aslin, R. N. (2007). Perceived object trajectories during occlusion constrain visual statistical learning. *Psychonomic Bulletin & Review*, 14(1), 173-178.
- Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review*, 16(3), 486-490.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49-71.
- Glennan, S. (2005). Modeling mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 375-388.

- Glennan, S. (2010). Mechanisms, Causes, and the Layered Model of the World. *Philosophy and Phenomenological Research*, 81(2), 362-381.
- Gopnik, A. & Schulz, L. (2007). *Causal Learning*. Oxford: Oxford University Press.
- Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: statistical learning in cotton-top tamarins. *Cognition*, 78, B53-B64.
- Illari, P. (2013). Mechanistic Explanation: Integrating the Ontic and Epistemic. *Erkenntnis*, 78, 237-255.
- Johnson, E. K. & Tyler, M. D. (2010). Testing the limits of statistical learning for word segmentation. *Developmental Science*, 13(2), 339-345.
- Kaplan, D. M. & Bechtel, W. (2011). Dynamical Models: An Alternative or Complement to Mechanistic Explanations? *Topics in Cognitive Science*, 3, 438-444.
- Kaplan, D. M. & Craver, C. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78(4), 601-627.
- Keidel, J. L., Jenison, R. L., Kluender, K. R., & Seidenberg, M. S. (2007). Does grammar constrain statistical learning? *Psychological Science*, 18(10), 922-923.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 83, B35-B42.
- Levy, A. (2013). Three kinds of new mechanism. *Biology & Philosophy*, 28, 99-114.
- Levy, A. (2013b). What was Hodgkin and Huxley's achievement? *The British Journal for the Philosophy of Science*, 0, 1-24.
- Levy, A. & Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of Science*, 80(2), 241-261.



- Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127-162.
- Newport, E. L., Hauser, M. D., Spaepen, G., & Aslin, R. N. (2004). Learning at a distance II. Statistical learning of non-adjacent dependencies in a non-human primate. *Cognitive Psychology*, 49(2), 85-117.
- Peña, M., Bonatti, L. L., Nespore, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.
- Peperkamp, S., Le Calvez, R., Nadal, J. P., & Dupoux, E. (2006). The acquisition of allophonic rules: statistical learning with linguistic constraints. *Cognition*, 101(3), B31-B41.
- Perruchet, P. & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *TRENDS in Cognitive Sciences*, 10(5), 233-238.
- Perruchet, P. & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, 39, 246-263.
- Saffran, J. R. (2001). Words in a sea of sounds: the output of infant statistical learning. *Cognition*, 81, 149-169.
- Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, 47(1), 172-196.
- Saffran, J. R. (2003). Statistical language learning: mechanisms and constraints. *Current Directions in Psychological Science*, 12(4), 110-114.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

- Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, 70, 27-52.
- Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: the role of distributional cues. *Journal of Memory and Language*, 35, 606-621.
- Saffran, J. R. & Thiessen, E. D. (2007). Domain general learning capacities. In E. Hoff & M. Schatz (ed.), *Blackwell Handbook of Language Development* (pp.68-86). Oxford: Blackwell Publishing.
- Saffran, J. R. & Wilson, D. P. (2003). From syllables to syntax: multilevel statistical learning by 12-month-old infants. *Infancy*, 4(2), 273-284.
- Sheredos, B., Burston, D. C., Abrahamsen, A., & Bechtel, W. (2013). Why do biologists use so many diagrams? *Philosophy of Science*, 80, 931-944.
- Thompson, S. P. & Newport, E. L. (2007). Statistical learning of syntax: the role of transitional probability. *Language Learning and Development*, 3(1), 1-42.
- Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, 97(2), B25-B34.
- Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(2), 399-407.
- Turk-Browne, N. B., Junge, J. A. & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, 134(4), 552-564.
- Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible visual statistical learning: transfer across space and time. *Journal of Experimental Psychology: Human Perception and Performance*, 35(1), 195-202.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning. *Journal of Cognitive Neuroscience*, 21(10), 1934-1945.

Turner, R. (2014). The philosophy of computer science. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*.

<http://plato.stanford.edu/archives/sum2014/entries/computer-science/>.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science*, 69, S366-S37.

Wright, C. D. (2012). Mechanistic explanation without the ontic conception. *European Journal for Philosophy of Science*, 2, 375-394.

