Mechanistic Explanation in Psychology[1]

Forthcoming in *The SAGE Handbook of Theoretical Psychology*. (Eds.) Hank Stam and Huib Looren de Jong.

Mark Povich

1. Introduction

Philosophers of psychology debate, among other things, which psychological models, if any, are (or provide) mechanistic explanations. This should seem a little strange given that there is rough[2] consensus on the following two claims: 1) a mechanism is an organized collection of entities and activities that produces, underlies, or maintains a phenomenon, and 2) a mechanistic explanation describes, represents, or provides information about the mechanism producing, underlying, or maintaining the phenomenon to be explained (i.e. the explanandum phenomenon) (Bechtel and Abrahamsen 2005; Craver 2007). If there is a rough consensus on what mechanisms are and that mechanistic explanations describe, represent, or provide information about them, then how is there no consensus on which psychological models are (or provide) mechanistic explanations? Surely the psychological models that are mechanistic explanations are the models that describe, represent, or provide information about mechanisms. That is true, of course; the trouble arises when determining what exactly that involves. Philosophical disagreement over which psychological models are mechanistic explanations is often disagreement about what it means to describe, represent, or provide information about a mechanism, among other things (Hochstein 2016; Levy 2013). In addition, one's position in this debate depends on a host of other seemingly arcane metaphysical issues, such as the nature of mechanisms, computational and functional properties (Piccinini 2016), and realization (Piccinini and Maley 2014), as well as the relation between models, methodologies, and explanations (Craver 2014; Levy 2013; Zednik 2015). Although I inevitably advocate a position, my primary aim in this chapter is to spell out all these relationships and canvas the positions that have been taken (or could be taken) with respect to mechanistic explanation in psychology, using dynamical systems models and cognitive models (or functional analyses[3]) as examples.

In Section 2, I lay out the basic conceptual toolkit of and motivation for a mechanistic account of explanation, including only recent historical development (for a more extensive history of mechanistic philosophy, see Chapters 2 and 3 of Glennan and Illari [2017]). In Section 3, I analyze more closely the question of what it takes for an explanation to be mechanistic. Taking center stage is an increasingly common distinction between mechanistic explanations, on the one hand, and their representational form, including the strategies and methodologies used to construct those representations, on the other (Andersen 2014a, b; Craver 2014; Craver and Kaplan 2018; Hochstein 2016; Levy 2013; Zednik 2011, 2015). I illustrate the way this distinction is used with regards to dynamical systems models, which dynamicists have claimed to be non-mechanistic explanations. A similar dialectic occurs with respect to the mechanistic status of functional analyses or cognitive models. I take this up in Section 4, where I examine the issue of the autonomy of psychology and the relation between functionalism and mechanistic explanation. In Section 5, I briefly compare the previous distinctions (between explanation and representational form/strategy) with the long-standing, though changing, distinction between ontic and epistemic conceptions of scientific explanation.

2. Mechanisms and Mechanistic Explanation

I first briefly gesture at an ontology of mechanisms, laying out only the bare commitments required to establish a broad concept of mechanism and mechanistic levels. Then, I motivate a mechanistic account of explanation, and make two normative distinctions: between mechanism schemata and mechanism sketches, and between how-possibly and how-actually models. I also contrast both of those distinctions with phenomenal models.

2.1 Mechanisms

While it is true that there is rough consensus that mechanisms contain entities and activities, or simply active entities, spatiotemporally organized to give rise to a behavior or property of the whole mechanism[4], there is disagreement over the specific metaphysics of mechanisms (see Illari and Williamson [2012] for a discussion of this disagreement and a recommendation of a broad construal of mechanisms, similar to mine, that applies across sciences). I do not wish to get involved in this debate here. I will assume a permissive[5] concept of mechanism as any collection of entities, also broadly construed, whose collective, organized activity gives rise to the behavior or property of a whole in context (see also Levy's [2014: 9] distinction between the 'narrow picture' and the 'broad picture' of mechanisms and Andersen's [2014a, b] distinction between mechanism_1 and mechanism_2). The entities in a mechanism need not be neatly localizable or contained within well-defined boundaries. An entity could be any set of structural properties that is robustly detectable (Piccinini and Craver 2011: 296).

Though permissive, this concept of mechanism is not trivial because it does not make every system – not even every *causal* system – a mechanism. Mechanisms contrast with aggregates, which lack the requisite organization. The parts of mechanisms have spatiotemporal properties, and stand in organizational and causal relations to one another, that are explanatorily relevant to the behavior of the mechanism as a whole. As such, mechanisms are more than the sums of their parts: their behavior depends on the spatial, temporal, and causal organization of their parts. Aggregates, in contrast, are systems – even causal systems – whose behavior does not depend on the spatial, temporal, and causal organization of their parts. As such, a property of an aggregate is literally a sum of the properties of its parts. The concentration of a fluid, for example, is an aggregation of particles. Aggregates have properties that do not change when their parts are reorganized, because in true aggregates, spatial, temporal, and causal organization is irrelevant (Wimsatt 1997; Povich and Craver 2017).

Mechanisms are often organized hierarchically into levels (Craver 2015; Povich and Craver 2017). The components of mechanisms can themselves be composed of organized components that are responsible for their activity. Similarly, a mechanism may compose an active entity that is itself a component in a larger mechanism. The term 'mechanistic levels' refers to this embedded, hierarchical organization of mechanisms.

Mechanistic levels contrast with another prominent use of the term 'levels' in psychology: Marr's levels or Marrian levels (Marr 1982). Marr's levels are best understood as levels of description, abstraction, analysis, or realization. The computational level, algorithmic level, and implementational level arguably do not stand in causal or componency relations with one another (Craver 2015; Craver and Bechtel 2007) (I briefly return to this in Section 4).

Mechanistic levels are necessarily local, in contrast to the more monolithic levels of Oppenheim and Putnam (1958), who divided nature into levels of atoms, molecules, cells, organs, organisms, and societies. For mechanistic levels, an entity is at a lower mechanistic level than another if and only if it is a component in the mechanism of the latter. From this, a weak notion of sameness of level is derived: two entities are at the same level only if they are components in the same mechanism, and neither is a component of the other.[6]

A component of a mechanism is more than just a mereological part; it is a part that contributes to the behavior of the mechanism – it is a constitutively *relevant* part. There is some debate over how to cash out this notion of constitutive relevance. Craver (2007) characterizes it in terms of mutual manipulability of part and whole: A part is a component of (or is constitutively relevant to) a mechanism if one can manipulate the mechanism’s behavior by manipulating the part, and one can manipulate the part’s behavior by manipulating the mechanism. This account is not without problems[7], but I will not examine those here. Instead, I will assume that the notion of constitutive relevance as contribution to the behavior of a mechanism is clear enough for our purposes.

2.2 Mechanistic Explanation

The contemporary account of mechanistic explanation has its origin primarily in the work of Salmon (1984, 1989), among others (see, for example, Scriven [1959, 1975]). He developed a causal account of explanation in response to problems that arose for the deductive-nomological account (DN;

also known as the covering-law model, sometimes as one half of the covering-law model). According to DN, an explanation is a deductive argument with descriptions of at least one law of nature and antecedent conditions as premises and a description of the explanandum phenomenon as the conclusion (Hempel and Oppenheim 1948). To explain, then, is to show that the explanandum phenomenon is *expectable* or *predictable* given the truth of the premises. However, tying explanation this closely to prediction generates some now-infamous problems (Salmon 1989). For example, on such an account, many mere correlations come out as explanatory, which intuitively is not true. A falling barometer reliably predicts the weather, but the falling barometer does not explain the weather (Salmon 1989: 47).

According to Salmon's (1984) causal-mechanical view, in contrast, explanation involves “situating” the explanandum phenomenon in the causal structure of the world. (Salmon called this an 'ontic conception' of scientific explanation, contrasting it with the 'epistemic conception' of the deductive-nomological account. I return to this still-relevant distinction in Section 5.) There are several ways of so situating an explanandum phenomenon. An etiological-causal explanation is 'backward-looking': it describes the explanandum phenomenon's past causal history (its immediately prior causes, the causes of those causes, and so on). A constitutive-mechanistic explanation is 'downward-looking': it describes the entities, activities, and organization of the mechanism that produces, underlies, or maintains the explanandum phenomenon. It is the kind of explanation that most readily comes to mind when one hears the phrase 'mechanistic explanation.' However, there is also the neglected contextual-mechanistic explanation, which is 'upward-looking' (though see Craver [2001] and Bechtel [2011], from which I have borrowed the 'looking' metaphor). It describes the broader environmental conditions on which the behavior of a mechanism depends, such as the character of its inputs[8]. For example, Bechtel (2011) considers Gibson’s (1979) ecological psychology to provide explanations of this sort for visual perception. This chapter will be concerned with the latter two kinds of explanation, constitutive and contextual mechanistic explanation.

There are two important normative distinctions (or continua) in the mechanist's conceptual framework: mechanism schemata versus mechanism sketches, and how-possibly versus how-actually models (Machamer, Darden, and Craver 2000; Craver 2007). A mechanism schema is an abstract description of a *type* of mechanism, rather than a specific token instance. Details will inevitably be omitted, but, ideally, only details that are irrelevant to the mechanism type. Details that are specific to tokens of the type can be added as the schema is applied to instances (Machamer et al. 2000: 15). Mechanism sketches, on the other hand, are incomplete descriptions of (type or token) mechanisms that contain black boxes and filler terms (Craver 2007: 113). They are still partially explanatory, but they are lacking in relevant detail. More details can be added to the model to fill in the gaps, though no model is ever fully complete, just complete enough for practical purposes (Craver and Darden 2013). Idealized models qualify as mechanism schemata, rather than sketches, to the extent that they capture relevant aspects of mechanisms.

A how-possibly model describes a merely possible mechanism, whereas a how-actually model describes the mechanism that (we have the most evidence to believe) actually produces, maintains, or underlies the explanandum phenomenon. This distinction is epistemic: turning a how-possibly model into a how-actually model does not require modifying the model itself in any way; it requires testing the model (Weiskopf 2011). The greater the evidential support for the model, the more how-actually it is. Between how-possibly and how-actually models is a range of how-plausibly models. Turning a mechanism sketch into a more complete mechanism schema, in contrast, requires modifying the model by filling in missing details (Craver and Darden 2013). These details may be at the same mechanistic level as the rest of the details in the model, or they may be at a lower mechanistic level.

In contrast to how-possibly and how-actually models, and mechanism sketches and schemata, which more or less completely describe possible or actual mechanisms responsible for some explanandum phenomenon, a merely descriptive, or phenomenal, model merely describes an explanandum phenomenon, usually in a general, concise way. Snell's law is a common example of a phenomenal model (Craver and Darden 2013). It accurately and compactly describes the relationship between the angle of incidence and the angle of refraction when light passes between two media, but it does not explain refraction. Somewhat like sketches, Glennan (2017) also includes how-roughly models, which only roughly accurately represent an actual mechanism.

Mechanistic explanations satisfy what are widely considered, by mechanists and non-mechanists (e.g., Chirimuuta 2014; Rice 2015; Weiskopf 2011) alike, to be important normative constraints on explanation: the ability to answer relevant counterfactual questions about the explanandum phenomenon ('what-if-things-had-been-different' questions or, more compactly, w-questions), and the ability to manipulate and control the explanandum phenomenon (Craver 2007)[9]. These norms capture in part what is distinctive about the scientific achievement of explanation, as opposed to other achievements like prediction, description, or categorization. As the barometer example above shows, a model can be predictive without being explanatory. These norms also provide a basis for explanatory power: when all else is equal, a model is more explanatorily powerful when and only when it can answer more relevant w-questions and afford more opportunities for control (Ylikoski and Kuorikoski 2010).

3. Dynamical Models, Strategies, and Explanations

I have briefly described what mechanisms and mechanistic explanations are, but I have not yet given any examples of explanations in psychology that are uncontroversially mechanistic. I will briefly describe a favorite of mechanists (see, e.g., Craver, 2005, 2007): the learning and memory mechanism of long-term potentiation (LTP; see Kandel, Schwartz, and Jessell 2000). There are different kinds of LTP, but I will focus on the commonly discussed NMDA-receptor dependent LTP. The entities in the mechanism include glutamate molecules, NMDA and AMPA receptors, and calcium and magnesium ions, and they engage in organized activities such as diffusing, blocking, opening, and binding, in order to produce long-term strengthening of synaptic connections. Specifically, glutamate is released from the presynaptic cell with each action potential and binds to NMDA receptors on the postsynaptic cell, causing the receptors to change shape, exposing a channel in the cell membrane. However, the channel is blocked by a magnesium ion as long as the postsynaptic cell is polarized (i.e., inactive). If the postsynaptic cell depolarizes as a result of frequent action potentials stimulating the AMPA receptors, the magnesium ion is released, allowing calcium ions to diffuse into the cell. The rising intracellular calcium concentrations set in motion a long biochemical cascade that results in the production of more AMPA receptors, eventually strengthening the synaptic connection.

Canonical examples of mechanistic explanation, such as the one I have just described, have given the impression that a mechanistic explanation should look a certain way or be constructed using certain methods (Zednik 2011; 2015). In some of the most seminal work on mechanistic explanation (e.g., Bechtel and Richardson 1993; Glennan 1996; Machamer et al. 2000), the examples and diagrams used were very machine-like: biological oxidation, voltage-gated ion channels, the action potential, protein synthesis, LTP. This arguably led to the impression that a mechanistic explanation is a particular machine-like kind of model or representation (Hochstein 2016; Zednik 2015), but some mechanists deny this (Craver 2014; Piccinini and Craver 2011; Zednik 2011, 2015).[10] With this impression in place, counterexamples to mechanistic explanation have come in the form of explanatory models in psychology (and elsewhere) that look nothing like the mechanists' canonical, machine-like examples. Implicit or explicit in many mechanists' responses to these counterexamples is a distinction between mechanistic explanations and mechanistic models, including the representational form that such explanations take[11]. Let us examine in some detail the dialectic in one prominent case from psychology – dynamical systems models – with that distinction in mind.

3.1 Dynamical Systems Models

Dynamical systems models are models that employ the mathematical concepts of dynamical systems theory, such as differential or difference equations (Chemero 2009; Izhikevich 2007; Zednik 2011). This allows the modeling of the temporal evolution of relevant variables, which can be represented mathematically (and graphically) as a trajectory through a phase or state space. The state space of a system represents all its possible states (i.e. all possible values of the system's variables). A trajectory through state space is thus a representation of how the system's variables change over time. Graphical representations have the benefit of allowing careful and intuitive analysis of state space topology, revealing abstract, dynamical features such as the presence of attractors (i.e. states into which the system tends from surrounding states) (Izhikevich 2007). In dynamical models in psychology, the relevant variables often span brain, body, and environment (van Gelder 1998; van Gelder and Port 1995; Zednik 2011). I briefly describe three dynamical models: the HKB model, Beer's model of categorical perception, and Thelen et al.'s (Thelen, Schöner, Scheier, and Smith 2001; Smith and Thelen 2003) dynamical field model of the A-not-B error.[12]

One of the first dynamical models that was presented as a challenge to mechanistic explanation was the Haken-Kelso-Bunz model (HKB; Haken, Kelso, and Bunz 1985. See also Chemero 2009; Chemero and Silberstein 2011; Stepp, Chemero, and Turvey 2011; Walmsley 2008). Although the explanandum of this model is not an especially cognitive phenomenon, it will be helpful to review it and mechanists' responses.

HKB is a model of bimanual coordination, specifically simultaneous, side-to-side movement of the index fingers (and hands). The behavioral data were obtained by asking participants to move horizontally both index fingers either in-phase (pointing toward the midline, then away) or out-of-phase (both pointing left, then both right). Participants were asked to keep pace with a metronome so that experimenters could manipulate the rate of finger movement (Kelso 1981). By increasing the rate, experimenters found that only in-phase movement is possible beyond a certain critical rate. Participants who began out-of-phase involuntarily switched to in-phase once the critical rate was crossed. The same phenomenon occurs during other forms of bimanual coordination, such as hand movements at the wrist (Kelso 1984).

To model this phenomenon with dynamical systems theory, the fingers are represented as coupled oscillators and the stable in-phase and out-of-phase movements as attractors. The dynamics are described by the following differential equation:

$$d\phi/dt = -dV/d\phi = -a \sin \phi - 2b \sin 2\phi,$$

where V is the so-called potential function, $V(\phi) = -a \cos \phi - b \cos 2\phi$, and the ratio b/a is a control parameter that varies inversely with finger oscillation frequency and determines the topology of the phase space (i.e. the landscape of attractors). At a low oscillation frequency, there are two attractors, corresponding to stable in-phase and out-of-phase movement. At a high frequency, past the critical value, the landscape shifts to include only one attractor, corresponding to stable in-phase movement. This accurately describes the observed behavioral data.

Beer's (1996, 2003) dynamical model of perceptual categorization (or categorical perception) is more cognitively interesting (Zednik [2011] provides a detailed analysis of this model). The model is a simulated system consisting of a 14-neuron continuous-time recurrent neural network (CTRNN), inside an evolved model agent (meaning its network architecture was constructed with an evolutionary algorithm[13]), inside a two-dimensional environment. The agent moves horizontally as circles or diamonds fall from above. It 'categorizes' these objects by catching the former and avoiding the latter. The agent perceives with an eye consisting of seven rays, each connected to a corresponding sensory input neuron. When a ray hits an object, its input neuron receives a signal inversely proportional to the distance from the object – the closer the object when 'seen' by a ray, the greater its input signal.

The agent with the best performance evolved a strategy of active scanning (Beer 2003). First, the agent centers the object in its field of view, then it moves back and forth, scanning the object. The scan narrows to home in on circles, while breaking to avoid diamonds. Beer (2003: 228-9) explains this active scanning as follows. First, he decomposes the agent-environment dynamics into the effect of the relative positions of agent and object on the agent's motion, and vice versa. Then, for both circle and diamond trials, he superimposes the motion trajectory of the object through the agent's field of view onto a steady-state velocity field, which represents, for each point in the agent's field of view, the agent's steady-state horizontal velocity in response to an object at that point (228). Finally, he notices from an examination of the agent's motion trajectories that it consistently overshoots the midline of its visual field, due to the lag in time for the neural network to respond to sudden changes in sensory input. Therefore, according to Beer, active scanning is explained by the dynamic interaction of the steady-state velocity fields and the neural network's lag.[14]

Finally, consider Thelen et al.'s (2001; Smith and Thelen 2003) dynamical field model of the A-not-B error. The A-not-B error is an instance of perseverative reaching – infants between 7-12 months will continue to reach to a location (location "A") where they have previously reached for a hidden toy, even after they see the toy moved to a new location (location "B"). Infants older than 12 months tend not to make the error. Getting into all the mathematical details of the dynamical field model of the A-not-B error would take us too far afield, so I will only focus on the broad outline. The model consists of a differential equation that specifies the activation level for each point in the infant's movement planning field. These points correspond to points in the infant's visual field where it could reach, it's reaching field. Once activation at a point of the movement planning field passes a threshold, the infant reaches toward the corresponding point in space. The activation is a function of three inputs: task input (e.g., environmental features like distance to containers and their salience), specific input (e.g., to where the experimenter draws the infant's attention) and memory input (reflecting previous reaching trials). The A-not-B error occurs when, after several A-trials, the memory input overwhelms the other inputs and the infant reaches to location A. This model successfully predicts many facts about the circumstances under which the A-not-B error occurs, such as the influence of posture, attention, and delay (Smith and Thelen 2003).

Dynamicists have argued that dynamical models such as the above are non-mechanistic because they abstract from low-level neural details and capture high-level qualitative behavior, yet such models are still explanatory because they yield understanding (Gervais 2015), accurate prediction, and unification of diverse systems (Chemero 2009; Chemero and Silberstein 2011; Stepp, Chemero, and Turvey 2011; van Gelder and Port 1995; Walmsley 2008). HKB, for example, does not include any specification of the neural mechanisms responsible for finger movements, but it does accurately describe diverse systems (including the coordinated limb movements of two separate people [Schmidt, Carello, and Turvey 1990]) and accurately predicts the amount of time it takes for the relative phase to stabilize following selective interference (Walmsley 2008). Van Eck (2018) disagrees that understanding, prediction, and unification are (individually or jointly) sufficient for explanation but agrees that some dynamical models – such as the dynamical field model – are not mechanistic. Instead, the dynamical field model is what he calls a “causal contextualized model” that does not refer to any constitutively relevant parts of a mechanism.

3.2 Mechanist Responses

The responses of mechanists to dynamical models have invoked the distinction laid out in Section 2.2 between predictive, phenomenal models and mechanism-schemata. Mechanists have argued that although a dynamical model's predictive power is a theoretical virtue, it is not enough to make a dynamical model explanatory (as the barometer example above shows), and, relatedly, that a dynamical model's unificatory ability (i.e. its ability to apply to a wide range of diverse systems) is likewise insufficient for explanation (Kaplan and Craver 2011). Instead, according to mechanists, a model like HKB, insofar as any internal causal structure is omitted[15], is a phenomenal model that merely describes an interesting, widespread pattern, but does not explain that pattern. In light of these concerns, Kaplan and Craver (2011) have argued that dynamicists have not yet provided a satisfactory account of what makes dynamical models explanatory, if they do not refer in any way to mechanisms or their organization (see also Kaplan 2015; Kaplan and Bechtel 2011).

Beer (2003) seems not to have explicitly taken his dynamical explanation to be non-mechanistic (see endnote 14). Zednik (2011) argues that Beer's explanation should be seen as describing interactive components in an extended mechanism that spans brain, body, and environment. The explanandum is the behavior of one component in this mechanism, the agent's active scanning. The model shows how interactions with the environment, along with the time lag in responding to stimuli, result in active scanning.[16] While this explanation does not describe any internal mechanisms, so is not a constitutive mechanistic explanation, it does seem to qualify as a kind of contextual mechanistic explanation.[17]

Regarding the dynamical field model (Thelen et al. 2001; Smith and Thelen 2003), let us consider van Eck's argument more closely. By "mechanistic explanation," van Eck explicitly has in mind constitutive explanation (2018: 14-5). Since the dynamical field model of the A-not-B error does not refer to a mechanism (or components thereof) inside the infant, but variables outside (e.g., posture and the locations of the containers), it is not a constitutive explanation. The mechanist is likely to concede this though and respond that it is still a mechanistic explanation, just of the contextual variety (Zednik 2011). We have already seen that Zednik (2011) considers Beer's model an extended (i.e., contextual) mechanistic explanation, and in the same paper Zednik argues that Thelen et al.'s model is too, for the same basic reasons. The dynamical field model is explanatorily similar to Beer's model, so van Eck would likely also consider Beer's model a causal contextualized model.

Van Eck considers this response (2018: 15-6) and argues that the dynamical field model is not a contextual mechanistic explanation because variables like posture are not causes but constraints that set the context for causes, constraints which, importantly, he says are different from background conditions. The mechanist is likely simply to deny either the cause/constraint distinction or the constraint/background condition distinction (as Gervais and Weber [2011], Zednik [2011], and even

Smith and Thelen [2003] appear to). After all, the mechanist will say, the input variables of the model, like posture or container color, stand in manipulable counterfactual dependence relations to the explanandum, and that is all it takes to be a cause of the explanandum (Woodward 2003). This is consistent with Bechtel's (2009: 557-9) discussion of how contextual mechanistic explanations describe the specific character of a mechanism's inputs and the systematic dependence of its behavior on them.

3.3 Strategies and Explanations

In responding to dynamicists, Zednik (2011) makes an increasingly common distinction between mechanistic explanations, on the one hand, and the heuristic strategies and tools used for constructing and representing them, on the other. He emphasizes that dynamical systems theory is a mathematical and conceptual framework that, as such, can be used to represent anything to which its concepts apply. If that includes the components, activities, and organization of mechanisms, then dynamical systems theory can provide mechanistic explanations.[18] Zednik (2015) has since extended this point, using examples from evolutionary robotics and network science to show how new tools for mechanism description and discovery go beyond the traditional mechanistic strategies of decomposition and localization (Bechtel and Richardson 1993).

Hochstein (2016) hits on a distinction similar to Zednik's (2015) in his diagnosis of the disagreement over which models are mechanistic explanations. He locates two opposing assumptions concerning the role of representation in mechanistic explanation. He calls these assumptions the 'representation-of' and 'representation-as' accounts of mechanistic explanation. According to the representation-of account, for an explanation to be mechanistic, it must be a representation of a mechanism, where this requires only that the explanation provide information about a relevant mechanism. According to the representation-as account, for an explanation to be mechanistic, it must not only provide information about a mechanism, but also represent the mechanism mechanistically, that is, as a mechanism. That is, not only must the represented thing in the world be a mechanism, it must be represented in a particular, machine-like way; the model or representation itself must have the form of depictions of neatly localized entities interacting to produce the explanandum.[19]

On the representation-of account, the general relation between mechanistic explanations and models is as follows. In the world there is a target mechanism, that produces, maintains, or underlies an explanandum phenomenon. There are many, conceptually and representationally distinct ways of describing this mechanism. To the extent that a model accurately picks out the mechanistic structures relevant to the explanandum phenomenon, the model mechanistically explains the explanandum phenomenon, regardless of how those structures are represented and regardless of which concepts are deployed (see Potochnik [2016] for some contrasting arguments). The form of representation and concepts deployed may be important when we are concerned with the understanding a model provides to cognitive agents, but mechanists who hold a view like the representation-of account typically also hold, contra Gervais (2015)[20], that explanation and understanding should be kept relatively distinct – a model that provides understanding but no information about mechanisms is never a mechanistic explanation (e.g., Craver 2014).

The representation-of account places no requirements on the form of the representation or concepts deployed. Since the representation-as account requires more of a model for it to be a mechanistic explanation, fewer psychological models will be counted as mechanistic explanations according to it than according to the representation-of account. Here we see, then, how the two opposing assumptions lead to disagreement about which psychological models are mechanistic and why.

Hochstein (2016) argues that the representation-of account trivializes the claim that neuroscience provides mechanistic explanations. Since the brain is a collection of mechanisms and neuroscientists model the brain, they therefore provide mechanistic explanations. However, some have argued that there are explanations in neuroscience that are not of mechanisms (e.g., Chirimuuta 2014; see Paz [2017] for a response). Furthermore, showing why and how neuroscientists provide mechanistic explanations requires showing why and how their distinctive concepts provide information about mechanisms, which is a controversial and nontrivial philosophical task, especially for cognitive, computational, and systems neuroscience (Piccinini and Craver 2011, Kaplan 2011, Zednik 2015, Povich 2015, 2019; I return to this in the next section). Compare this to a similar debate about etiological causal explanation. For example, Skow (2014) basically holds a representation-of account of *causal* explanation: roughly, an explanation is causal if and only if it provides information about the explanandum's causal history.[21] One might say that this trivializes the claim that science provides causal explanations, since scientists model the natural world and, plausibly, the natural world is made of causes (ignoring the quantum realm and general skepticism about causation; see Andersen 2016). What is non-trivial is showing *how* a model provides a causal explanation. Skow responds to some prominent putative counterexamples to causal explanation (e.g., explanations that cite causally inert entities) by showing in detail precisely how they provide causal information (e.g., by showing how such explanations rule out possible causal histories). Furthermore, constitutive mechanistic models can provide information about the constitutive relation between an explanandum and its mechanism without providing much, if any, information about its causal history. These would not be causal explanations according to an account like Skow’s, so it does not make all scientific explanations causal (see Povich [2018] for a discussion of some explanations that are arguably non-causal even on a broad, informational construal of causal explanation).

The representation-of account of mechanistic explanation is therefore not without some precedent. An account like Skow's (2014) has long been widely recognized as legitimate in the literature on causal explanation, where, to be a causal explanation, a representation need only provide information about an explanandum's causes or causal history (Jackson and Pettit 1990; Lange 2013;

Lewis 1986). On such accounts of causal explanation, explanations do not have to have a particular representational or conceptual form in order to count as causal. Proponents of the representation-of account can be seen as extending this idea to mechanistic explanation.

A distinction similar to Zednik's (2011; 2015) is made by Levy (2013), who distinguishes between what he calls 'causal mechanism,' 'explanatory mechanism,' and 'strategic mechanism'. Only the latter two concern us here. According to Levy, explanatory mechanism is the thesis that 'to explain a phenomenon, one must cite mechanistic information' (100). This appears to be equivalent to the representation-of account. On the other hand, strategic mechanism 'articulates a way of doing science, a framework for representing and reasoning about complex systems,' using modeling methods such as decomposition and localization (104-5). Unlike the representation-as account, strategic mechanism does not explicitly say that mechanistic explanations must have a certain representational form, but such strategies do constrain the representational form of models. Adherence to strategic mechanism might therefore motivate adherence to the representation-as account.

Similarly, Andersen (2014a, b) distinguishes between five conceptions of mechanism or mechanistic philosophy. The most important for us are what she calls mechanism_1 and mechanism_2. The central difference between them is that mechanism_1 has stricter criteria (like a regularity requirement) for when something counts as a mechanism and, like Levy's (2013) 'strategic mechanism,' offers methodological prescriptions while mechanism_2 does neither. Like Levy's (2013) 'explanatory mechanism,' according to mechanism_2, models are explanatory because they describe mechanisms in the permissive sense (Andersen 2014a: 280).

Thus, Hochstein's (2016) representation-of/representation-as distinction, Zednik's (2015) explanation/heuristic distinction, Levy's (2013) explanatory mechanism/strategic mechanism distinction, and Andersen's (2014a) mechanism_2/mechanism_1 distinction seem to be at least *roughly* coextensive. One might also make a similar point by distinguishing between mechanistic explanations and mechanistic models, construed as kinds of representation. Mechanistic explanations can be provided by non-mechanistic models, since non-mechanistic models can provide information about mechanisms. It is important to note, however, that agreement on one side of these distinctions does not guarantee agreement about which psychological models are mechanistic explanations. For example, two philosophers who both hold the representation-of account could still disagree about which psychological models are mechanistic explanations, because they disagree, *inter alia*, about the ontology of mechanisms – that is, they disagree about whether the thing about which a model provides information counts as a mechanism (see endnote 5).

4. Abstraction, Functionalism, and Realization

In addition to dynamical models, functional analyses or cognitive models are prominent putative counterexamples to mechanistic explanation in psychology (Fodor 1965, 1968; see Piccinini and Craver [2011] for response). A functional analysis of a psychological capacity explains it in terms of the functional properties, either of the whole cognitive system, or of its parts. Contrast functional analysis with the explanation of LTP above: in that explanation, explicit reference is made to neatly localizable structural components (e.g., glutamate molecules, NMDA and AMPA receptors, and calcium and magnesium ions) that engage in organized activities (e.g., diffusing, blocking, opening, and binding) that are responsible for the explanandum phenomenon. Functional analysis, on the other hand, proceeds relatively independently (or autonomously) of consideration of the structural components (and their physical activities) that realize the functional properties or play the functional roles posited in the analysis (Weiskopf 2011; 2017). Mechanists have either denied such independence or denied that such independence renders functional analyses non-mechanistic, arguing instead that functional analyses are mechanism-sketches (call this the “sketch thesis”; Piccinini and Craver 2011; Piccinini 2015; Povich 2015). Let us examine more closely the reasons for and against the mechanistic status of functional analyses, which will bring out how realization and abstraction (can) relate to mechanistic explanation.

4.1 Functional Analyses and Mechanism Sketches

The primary reason that Piccinini and Craver (2011) give for the sketch thesis is that functional analyses put constraints on the possible mechanisms that implement the functions posited in the analysis. Similarly, structure constrains function: not just any structural component can perform any function. For example, to perform the functions of belief and desire boxes, a mechanism(s) must be able to distinguish between those two types of representation and transform them in relevant ways (Piccinini and Craver 2011: 303). This puts some constraints on what could possibly implement belief and desire boxes. Since putting constraints on a possible mechanism is the same as providing information about a mechanism (in the most common sense of "information"), this argument appears to rely on a representation-of account of mechanistic explanation (Hochstein 2016): functional analyses are mechanism sketches because they provide information about mechanisms.

The neural mechanisms that play the functional roles of belief and desire boxes (or other functional states), are likely vague, widely distributed, and multi-functional. For this reason, Piccinini and Craver (2011) also emphasize a permissive concept of mechanism like the one given in Section 2.1 above, according to which the components that play the functional roles need not be neatly localizable or contained within well-defined boundaries (Piccinini and Craver 2011: 296).

Weiskopf (2011; 2017) objects to the sketch thesis and the claim, required for that argument, that mechanism components can be widely distributed. Against the latter claim, he argues that it results in 'greater difficulty in locating the boundaries of mechanisms' (Weiskopf 2011: 315) and gives up 'any requirement that parts be describable in a way that our modeling techniques can capture' (2017: 56). I do not have space to respond in detail here, but I have argued in depth elsewhere that model-based fMRI can ameliorate these practical worries (Povich 2015).[22]

In response to the sketch thesis, Weiskopf (2017) argues as follows. If functional analyses are mechanism-sketches, then they are amenable to two kinds of elaboration. 1) Intralevel elaboration involves adding details, discharging filler terms, and so on, while staying at the same mechanistic level. 2) Interlevel elaboration involves going down mechanistic levels in order to explain their component entities and activities. He argues that it cannot be the case that functional analyses need interlevel elaboration in order to describe completely the causal structure relevant to a psychological phenomenon, because that would lead to a downward regress. In order to provide a complete model at any mechanistic level, one would have to give a complete model at every lower mechanistic level. He argues that if functional analyses need intralevel elaboration, this can be accomplished with more specific functional analyses of subsystems – there is no reason to think functional concepts can never fully accurately capture the psychological properties of a system.

The mechanist can make several moves in response to this argument. First, it could be argued that even if a functional analysis fully captures that psychological properties of a system, its explanatory incompleteness is shown by the fact that adding implementation details increases the explanatory power of the model (i.e., its ability to answer w-questions and to afford opportunities for intervention and control). Note that adding implementational details is not a kind of interlevel elaboration: simply to identify the occupant of a functional role is not to explain mechanistically how that occupant plays its role[23]. Implementational details are mechanistically intralevel details. To think otherwise is to confuse mechanistic levels with Marrian levels, as I have argued elsewhere (Povich 2019). Weiskopf appears to be confusing the two when he says, in a discussion about descending mechanistic levels, “What does not follow is that an explanation of a psychological capacity by appeal to a cognitive model also requires that we have a further set of *lower-level explanations* for how all of the elements of the model are *implemented*” (2017: 59; emphasis added).

Second, building off the previous point, the mechanist could accept that functional analyses can be complete mechanism schemata, rather than mere sketches. This seems a natural move for a proponent of the representation-of account, according to which the kinds of concept deployed in an explanation do not affect its mechanistic status. The mechanist could argue that as long as the functional concepts deployed pick out features of mechanisms, functional analyses count as mechanistic explanations. This is just to deny the representation-as account that is presupposed in Weiskopf's (2017) argument, for example, when he claims that, 'The question is whether remedying this sketchiness requires stepping out of the explanatory framework of psychology' (58) or that, 'An ideally complete cognitive model will still be one that is couched in the autonomous theoretical vocabulary of psychology' (59). For Weiskopf, whether an explanation is mechanistic depends on the vocabulary in which it is couched. Proponents of the representation-of account deny this.

Piccinini and Craver's (2011) argument that functional analyses are mechanism-sketches was not meant to imply that functional or computational analyses are never true or explanatory. The argument was that functional analyses are true and explanatory to the extent that they accurately describe mechanisms. A central part of that argument was showing that different kinds of functional analysis are 'elliptical' descriptions of mechanisms. This argument gains plausibility from a particular account of the realization relation, namely the subset account (Heil 2011; Shoemaker 2007; Piccinini and Maley 2014; Povich 2019). On the subset account, the realized is literally a subset of the realizer – functional concepts pick out subsets of the very same properties picked out by non-functional (structural, neural) concepts. Thus, realized and realizer seem not to be as ontologically autonomous as a functionalist like Weiskopf needs them to be. The subset account of realization, then, gives more precise content to the claim that functional analyses are 'elliptical' descriptions of mechanisms. The subset account also makes explicit why adding implementational details to a functional analysis is not a mechanistically interlevel elaboration: to add implementational details is to identify certain properties of the realizer that are not picked out by a functional concept, all of which properties are at the same mechanistic level (see Povich [2019] for an elaboration of this argument).[24]

If all this is right, however, then it seems that Piccinini and Craver (2011; also Piccinini 2015) were wrong that all functional analyses are mechanism-sketches – they seem to have the conceptual resources to say that sometimes functional analyses can be complete(-enough) mechanism schemata. Perhaps it is more consistent with their view to say that all functional analyses are how-possibly models[25]. Mechanistic details are often sought in order to confirm functional analyses (a point van Eck [2018] makes with respect to dynamical models), though evidence for functional analyses can come from other places too, such as behavioral interventions (Weiskopf 2017).

5. Ontic and Epistemic Conceptions of Explanation

The distinctions set out in Section 3 are somewhat related to, but not as coextensive with, another that is prominent in contemporary philosophy of explanation: Salmon's (1984, 1989) distinction between epistemic and ontic conceptions of scientific explanation. These conceptions were different accounts of what a scientific explanation aims to show of its explanandum phenomenon: that it is expected to occur and that it fits 'into a discernible pattern,' respectively (1984, 121)[26]. According to Salmon, the 'discernible pattern' into which an explanandum phenomenon is fit is structured by causal processes, interactions, and laws (1984: 132). Explaining is 'providing information about these patterns that reveals how the explanandum-events fit in' (1989, 121). Explanation, for Salmon, is not about nomic expectability or nomic necessity, but about fitting the explanandum into 'discernible patterns' and 'relationships that exist in the world' (1984, 121) (Povich 2018).[27]

The ontic-epistemic debate has shifted twice since Salmon (Illari 2013)[28]. Salmon framed the debate in terms of what explanations do. After Salmon, the debate was framed metaphysically, as a debate about what explanations are: The ontic conception was associated with the claim that scientific explanations are (almost always causal) dependence relations in the world; the epistemic conception became associated with claim that scientific explanations are epistemic states or representations.

The distinction has since shifted from a metaphysical distinction concerning what explanations *are* – representations or the things represented? – to one that focuses on explanatory demarcation and normative constraints on explanation (Illari 2013; Craver 2014, 2019; Povich 2018).[29] Craver writes that according to the ontic conception, 'in order to satisfy these two objectives [of explanatory demarcation and explanatory normativity], one must look beyond representational structures to the ontic structures in the world' (2014: 28). The idea is that attention to ontic structures, rather than representational form, is required to demarcate explanation from other scientific achievements, like prediction, and to distinguish good from bad explanations, how-possibly from how-actually explanations, and explanatorily relevant from irrelevant features (2014: 51).[30]

The ontic conception/epistemic conception distinction is not coextensive with the distinctions in Section 3. The other distinctions concern what makes a model mechanistic, while the ontic conception/epistemic conception distinction concerns what makes a model an explanation at all. It seems that one could consistently hold, for example, that representational form does not matter for whether a model is an explanation but does matter for whether it is a *mechanistic* explanation. It also seems that one could consistently hold, for example, that representational form matters for whether a model is an explanation but does not matter for whether it is a mechanistic explanation.

However, the ontic conception and the representation-of account (or corresponding sides of the other distinctions) tend to go together, and opponents of the representation-of account (or corresponding sides of the other distinctions) tend to be opponents of the ontic conception (Wright 2012) – e.g., Hochstein (2016) cites Craver’s (2014) exposition and defense of the ontic conception when he characterizes the representation-of account. I think it is intuitive why this is. The ontic conception says that the difference between a model that explains and one that does not depends on the ontic structures about which the model provides information. The distinction between different kinds of explanation can then be easily drawn by appeal to the kind of ontic structure about which information is provided (Craver 2014; Povich 2018, forthcoming: 33). So, if a model provides information about a mechanism, it is a mechanistic explanation; if it provides information about causes, it is a causal explanation; etc. It is a virtue of this account that it makes mixed kinds of explanation intelligible, explanations that are partly etiological, partly mechanistic, etc. One could think of the representation-of account as an ontic conception of mechanistic explanation.

6. Conclusion

Which psychological models are (or provide) mechanistic explanations? The conciliatory answer is, 'It depends.' It depends, for example, on whether one adopts a representation-of account or representation-as account of mechanistic explanation (or one side of any other of the roughly coextensive distinctions; Hochstein 2016; Andersen 2014a, b; Levy 2013; Zednik 2015). In some cases, such as when all parties agree about the counterfactual dependencies and that these dependencies are explanatorily relevant, disputes about whether a psychological model is a mechanistic explanation may be semantic.[31] Whether to call a model a “mechanistic explanation” would then largely depend on how you conceive the mechanistic project. If you conceive it as 'explanatory mechanists' (Levy 2013) tend to, as articulating a 'downward' way of causally situating an explanandum phenomenon that was neglected by Salmon and others who focused on 'backward' (etiological) causal explanation (Craver 2007: 8), then it becomes clearer why one might hold something like the representation-of account (Povich 2015). From this perspective, a categorization or typology of the diverse kinds of mathematical framework, representation, or model used in scientific explanatory practices might be useful for some purposes but does not seem to advance the classical project of a philosophical theory of scientific explanation, which is to provide conditions of explanatory demarcation and normativity (Craver 2014; again, see Potochnik [2016] for a contrast).

Endnotes

1. Special thanks to Carl Craver and Eric Hochstein for many hours of stimulating conversation on nearly all of the ideas touched upon in this chapter. Thanks also to Gualtiero Piccinini for discussion of

his work on functionalism and mechanism and to an anonymous reviewer for very helpful comments. Mistakes are inevitably my own.

2. See Section 2.1 below.

3. As I will use the terms here, not all functional analyses are cognitive models, but all cognitive models are functional analyses. See Section 4.

4. I will speak of the 'behavior' or 'property' of a whole mechanism as that for which the mechanism is responsible, but there is also disagreement about how metaphysically to characterize the phenomenon produced by a mechanism (Kaiser and Krickel 2016).

5. The more restrictive one makes the concept of mechanism (for example, by requiring modularity and stability [Woodward 2013], localizability [Weiskopf 2011], or regularity [Andersen 2012]), the correspondingly rarer mechanistic explanations will be. When presented with a putatively non-mechanistic explanation, one should always ask what concept of mechanism is in the background.

6. Eronen (2015) argues that this is so weak that it is tantamount to abandoning the idea of levels altogether.

7. For example, Craver's account requires that the interventions used to establish constitutive relevance are 'ideal,' which seems conceptually impossible (Baumgartner and Gebharter 2015; Couch 2011; Harinen 2014; Leuridan 2012; Prychitko 2021; Romero 2015). For Craver’s revised account, see (Craver, Glennan, and Povich forthcoming).

8. Bechtel (2011) also includes what he calls 'looking around,' which involves determining how the components of a mechanism are organized. I have subsumed this under constitutive mechanistic explanation.

9. These are related, of course. The latter ability is a practical analogue of the former.

10. Although not all representations are models, in this context I will use the terms 'representation' and 'model' synonymously to mean some kind of structure (e.g., a concrete replica, a mathematical

equation, a diagram, or a linguistic description) that is interpreted to represent a target system (Weisberg 2013). This terminological choice runs roughshod over Weisberg's distinction between models and model descriptions, but this should not affect the points that follow.

11. I do not mean to imply that all mechanists make this distinction or that no mechanist has meant a certain kind of model by 'mechanistic explanation,' just that, of those mechanists who dispute putative examples of non-mechanistic explanation, many of them have appealed to this distinction, or something like it. Kaplan's (2011) model-to-mechanism-mapping (3M) requirement might preclude him from making this distinction, or at least from saying that models that fail 3M, but still provide information about mechanisms, are mechanistic explanations. For example, you can provide information about mechanisms by saying what is *not* responsible for an explanandum (cf. Lewis 1986: 220), but this would violate 3M. This comes down to Kaplan's intended scope of 3M, something about which I will refrain from speculating here.

12. Since my theme is psychology, I leave aside dynamical models in neuroscience, though they too have been presented as counterexamples to mechanistic explanation. For example, see Ross (2015), which relies on Batterman and Rice's (2014) notion of a 'minimal model explanation.' The response to Batterman and Rice in Povich (2018) applies to Ross's argument as well.

13. Specifically, the connection weights, biases, time constants, and gain were evolved, but not the number of nodes (Beer 2003: 214).

14. Beer (2003) also analyzes the neural network, including individual neurons, and how it changes over the course of active scanning, thus providing a *multilevel* mechanistic explanation of categorical perception via active scanning. For brevity's sake, and due to the fact that this part of Beer's analysis is more clearly consistent with mechanistic explanation, I omit further discussion of this; see Zednik (2011) for more.

15. Kaplan and Craver (2011) note that Kelso and colleagues have not neglected to investigate the

neural mechanisms that generate the dynamics HKB describes.

16. Questions like, 'Why is the lag such and such amount of time?' require looking at the neural mechanisms of the agent. This does not detract from the contextual mechanistic explanation of active scanning – the lag time is simply a different explanandum. See Zednik (2011: 254).

17. Chemero (2009: xi, 85) argues that Gibson's (1979) ecological psychology provides a background theory unifying all dynamical modeling in psychology. As we have seen, however, Bechtel (2011) argues that Gibson's ecological psychology provides contextual mechanistic explanations.

18. Similar arguments have been made with respect to optimization theory and network and graph theory: they are formal frameworks that can be used to provide mechanistic explanations when they provide information about properties – usually organizational properties – of a mechanism (or components thereof) that are relevant to the explanandum (see, e.g., Levy and Bechtel 2013; Craver 2016).

19. One might also put this by saying there is a distinction between mechanistic models and models of mechanisms (Craver, personal communication).

20. The mechanist is likely also to object to Gervais's (2015: 61) restrictive account of mechanistic explanation, which, like Weiskopf's (2011, 2017), requires highly specific (i.e., non-abstract), neatly localizable entities. Gervais (2015: 61-2) also mistakenly infers, from Kaplan and Bechtel's (2011: 443) claim that Voss' (2000; Stepp et al. 2011) strong anticipation model of circadian rhythms is a how-possibly model, that it is not mechanistic. But a how-possibly model is a mechanistic model, just one with little to no evidential support. The mechanist may say that no how-possibly models are explanations (as van Eck [2018: 10] notes), but they need not say they are not mechanistic – they provide information about a possible mechanism.

21. Skow's (2014) account is more complicated than this, and it is limited to explanations of particular events.

22. Model-based fMRI is a neuroimaging method that combines psychological models with fMRI data, allowing cognitive neuroscientists to explore how the components of psychological models map onto distributed brain regions.

23. Endicott (2011) helpfully distinguishes between the 'what' and the 'how' of functional realization. He argues that the former is answered by the subset account, and the latter is answered by the dimensioned account (Gillett 2002a, b), which is basically a kind of constitutive explanation. Only the latter requires descending mechanistic levels; the former concerns implementation as I have been discussing it. Polger and Shapiro (2008) argue that dimensioned realization is just an account of composition, not the realization of functional properties, because functions are not composite objects whose properties depend on the properties of their organized parts. If Polger and Shapiro are right, then the subset account is (currently) the only game in town for the functionalist, and the autonomist seems to be in trouble.

24. Here I ignore certain debates around powers and their ontological and individuative relations to properties. For this reason and others (Audi 2012; Heil 2011), I prefer to characterize the subset account of realization in terms of concepts and what they pick out, rather than properties and the powers they confer or bestow on objects that possess them.

25. Some later work of Piccinini's might seem to suggest this. For example, "Much of Marr's work belongs in this how-possibly category. … It is no longer enough to simply home in on ways in which problems *might* be solved in the brain; contemporary cognitive neuroscience aims to understand how those problems are *actually* solved in the brain" (Boone and Piccinini 2016: 1520; original emphasis).

26. Salmon also includes the modal conception of scientific explanation, according to which an explanation shows that the explanandum had to occur.

27. This need not be construed solely causally, and Salmon did not think causation was essential to the ontic conception (Povich 2018).

28. The modal conception has fallen out of favor and was not included in later debates (but see Lange [2013] for a recent defense).

29. According to Illari (241), Craver holds that this has always been the core of the debate between the ontic conception and the epistemic conception.

30. Under this framing of the debate, Wright (2012) overemphasizes the role that lexical ambiguity of the term 'explanation' plays in the argument for the ontic conception. The argument for Craver's claims about explanatory demarcation and normativity does not require lexical ambiguity.

31. Here I am extending an idea from Woodward (2003: 86), who wrote in the context of etiological causal explanation: “if, as I assume is the case, we agree about all of the relevant counterfactuals, it is not clear that there is a further question over and above these about whether the sergeant's orders "really" cause the corporal to advance or not, unless that question is simply a (misleading) way of asking about the judgments that most people in fact endorse in such a case. This assessment seems particularly compelling to the extent that our concern is with causal explanation, for once we have been given information about the complete patterns of counterfactual dependence in the symmetric overdetermination and trumping cases as well as a description of the actual course of events, it appears that nothing has been left out that is relevant to understanding why matters transpired as they did.”

References

Andersen, Holly (2012) ‘The case for regularity in mechanistic causal explanation’, *Synthese*, 189(3), 415-432.

Andersen, Holly (2014a) ‘A field guide to mechanisms: Part I’, *Philosophy Compass*, 9(4), 274-283.

Andersen, Holly (2014b) ‘A field guide to mechanisms: Part II’, *Philosophy Compass*, 9(4), 284-293.

Andersen, Holly (2016) ‘Complements, not competitors: causal and mathematical explanations’, *The British Journal for the Philosophy of Science*, 69(2): 485-508.

Audi, Paul (2012) ‘Properties, powers, and the subset account of realization’, *Philosophy and*

Phenomenological Research, 84(3), 654–74.

Baumgartner, Michael and Gebharter, Alexander (2015) 'Constitutive relevance, mutual manipulability, and fat-handedness', *British Journal for the Philosophy of Science*.

Bechtel, William (2009) ‘Looking down, around, and up: Mechanistic explanation in psychology’, *Philosophical Psychology*, 22, 543–64.

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

Beer, Randall D. (1996) 'Toward the evolution of dynamical neural networks for minimally cognitive behavior', in P. Maes, M. Mataric, J. A. Meyer, J. Pollack, and S. Wilson (eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press. pp. 421–9.

Beer, Randall D. (2003) 'The dynamics of active categorical perception in an evolved model agent', *Adaptive Behavior*, 11(4): 209–43.

Beer, Randall D. and Williams, Paul L. (2015) 'Information processing and dynamics in minimally cognitive agents', *Cognitive Science*, 39(1): 1–38.

Boone, Worth and Piccinini, Gualtiero (2016) ‘The cognitive neuroscience revolution’, *Synthese*, 193(5): 1509–34.

Bromberger, Sylvain (1966) 'Why questions', in R. G. Colodney (ed.), *Mind and Cosmos.* Pittsburgh: University of Pittsburgh Press. pp. 86–111.

Chemero, Anthony (2009) *Radical Embodied Cognitive Science.* Cambridge, MA: MIT Press.

Chemero, Anthony and Silberstein, Michael (2008) 'After the philosophy of mind: replacing scholasticism with science', *Philosophy of Science*, 75(1): 1–27.

Chirimuuta, Mazviita (2014) 'Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience', *Synthese*, 191(2): 127–53.

Couch, Mark B. (2011) 'Mechanisms and constitutive relevance', *Synthese*, 183 (3):375–88.

Craver, Carl F. (2001) 'Role functions, mechanisms, and hierarchy', *Philosophy of Science*, 68(1): 53-74.

Craver, Carl F. (2005) ‘Beyond reduction: Mechanisms, multifield integration and the unity of neuroscience’, *Studies in History and Philosophy of Science: Part C. Studies in History and Philosophy of Biological and Biomedical Sciences*, 36: 373–395.

Craver, Carl F. (2006) 'When mechanistic models explain', *Synthese*, 153(3): 355–76.

Craver, Carl F. (2007) *Explaining the Brain*. Oxford: Oxford University Press.

Craver, Carl F. (2014) 'The ontic account of scientific explanation', in M. Kaiser, O. Scholz, D. Plenge and A. Hüttemann (eds.), *Explanation in the Special Sciences: The Case of Biology and History*. New York: Springer. pp. 27–54.

Craver, Carl F. (2016) ‘The explanatory power of network models’, *Philosophy of Science*, 83(5), 698-709.

Craver, Carl F. (2019) ‘Idealization and the ontic conception: A reply to Bokulich’, *The Monist*, 102(4): 525-530.

Craver, Carl F. and Bechtel, William (2007) 'Top-down causation without top-down causes', *Biology and Philosophy*, 22(4): 547–63.

Craver, Carl F. and Darden, Lindley (2013) *In Search of Mechanisms.* Chicago: University of Chicago Press.

Craver, Carl F. and Tabery, James (2016) 'Mechanisms in science', The Stanford Encyclopedia of Philosophy (Spring 2016 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2016/entries/science-mechanisms/>.

Craver, Carl F. and Kaplan, David M. (2018) ‘Are more details better? On the norms of completeness for mechanistic explanations’, *The British Journal for the Philosophy of Science*.

https://doi.org/10.1093/bjps/axy015

Craver, Carl F., Glennan, Stuart and Povich, Mark (forthcoming) ‘Constitutive relevance and mutual manipulability revisited’, *Synthese*.

Endicott, Ronald P. (2011) 'Flat versus dimensioned: The what and the how of functional realization', *Journal of Philosophical Research*, 36: 191–208.

Eronen, Markus I. (2015) 'Levels of organization: A deflationary account', *Biology and Philosophy*, 30(1): 39–58.

Gervais, Raoul (2015) ‘Mechanistic and non-mechanistic varieties of dynamical models in cognitive science: Explanatory power, understanding, and the ‘mere description’ worry’, *Synthese*, 192(1): 43-66.

Gervais, Raoul and Weber, Erik (2011) ‘The covering law model applied to dynamical cognitive science: A comment on Joel Walmsley’, *Minds and Machines*, 21(1): 33-39.

Gibson, James J. (1979) *The Ecological Approach to Visual Perception*. Boston: Houghton-Mifflin.

Gillett, Carl (2002a) ‘The dimensions of realization: A critique of the standard view’, *Analysis*, 62(276): 316–323.

Gillett, Carl (2002b) ‘The metaphysics of realization, multiple realizability, and the special sciences’, *The Journal of Philosophy*, 100(11): 591–603.

Glennan, Stuart (1996) 'Mechanisms and the nature of causation', *Erkenntnis*, 44(1): 49–71.

Glennan, Stuart (2017) *The New Mechanical Philosophy*. Oxford University Press.

Glennan, Stuart and Illari, Phyllis (eds.) (2017) *The Routledge Handbook of Mechanisms and Mechanical Philosophy.* Routledge.

Haken, Hermann, Kelso, J. A. Scott and Bunz, Heinz (1985) 'A theoretical model of phase transitions in human hand movements', *Biological Cybernetics*, 51(5): 347–56.

Harinen, Totte (2014) 'Mutual manipulability and causal inbetweenness', *Synthese*, 1–20.

Heil, John (2011) ‘Powers and the realization relation’, *The Monist*, 94(1): 34–53.

Hempel, Carl and Oppenheim, Paul (1948) 'Studies in the logic of explanation', *Philosophy of Science*, 15(2): 135–75.

Hochstein, Eric (2016) 'One mechanism, many models: A distributed theory of mechanistic explanation', *Synthese*, 193(5): 1387–407.

Illari, Phyllis (2013) 'Mechanistic explanation: Integrating the ontic and epistemic', *Erkenntnis*, 78(2): 237–55.

Illari, Phyllis and Williamson, Jon (2012) 'What is a mechanism? Thinking about mechanisms across the sciences', *European Journal for Philosophy of Science*, 2(1): 119–35.

Izhikevich, Eugene M. (2007) *Dynamical Systems in Neuroscience*. Cambridge, MA: MIT Press.

Kaiser, Marie I. and Krickel, Beate (2016) ‘The metaphysics of constitutive mechanistic phenomena’, *The British Journal for the Philosophy of Science*, 68(3), 745-779.

Kandel, Eric R., Schwartz, James H., and Jessell, Thomas M. (2000) *Principles of Neural Science*. New York: McGraw Hill.

Kaplan, David M. (2011) 'Explanation and description in computational neuroscience', *Synthese*, 183(3): 339–73.

Kaplan, David M. (2015) 'Moving parts: the natural alliance between dynamical and mechanistic modeling approaches', *Biology and Philosophy*, 30(6): 757–86

Kaplan, David M. and Bechtel, William (2011) 'Dynamical models: An alternative or complement to mechanistic explanations?', *Topics in Cognitive Science*, 3(2): 438–44.

Kaplan, David M. and Craver, Carl F. (2011) 'The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective', *Philosophy of Science*, 78(4): 601–27.

Kelso, J.A. Scott (1981) 'On the oscillatory basis of movement', *Bulletin of the Psychonomic Society*, 18: 63.

Kelso, J.A. Scott (1984) 'Phase transitions and critical behavior in human bimanual coordination'. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, 246(15): R1000–R1004.

Lange, Marc (2013) 'What makes a scientific explanation distinctively mathematical?', *British Journal for the Philosophy of Science*, 64(3): 485–511.

Leuridan, Bert (2012) 'Three problems for the mutual manipulability account of constitutive relevance in mechanisms', *The British Journal for the Philosophy of Science*, 63(2): 399–427.

Levy, Arnon (2013) 'Three kinds of new mechanism', *Biology and Philosophy*, 28(1): 99–114.

Levy, Arnon (2014) 'Machine-likeness and explanation by decomposition', *Philosophers' Imprint*, 14(6): 1–15.

Levy, Arnon and Bechtel, William (2013) 'Abstraction and the organization of mechanisms', *Philosophy of Science*, 80(2): 241–61.

Lewis, David (1986) 'Causal explanation', in David Lewis, *Philosophical Papers, Vol. 2.* New York: Oxford University Press. pp. 214–40.

Machamer, Peter K., Darden, Lindley, and Craver, Carl F. (2000) 'Thinking about mechanisms', *Philosophy of Science* 67(1): 1–25.

O'Doherty, John, Hampton, Alan, and Kim, Hackjin (2007) 'Model-based fMRI and its application to reward learning and decision making', *Annals of the New York Academy of Sciences*, 1104(1): 35–53.

Paz, Abel Wajnerman (2017) ‘A mechanistic perspective on canonical neural computation’, *Philosophical Psychology*, 30(3): 209-230.

Piccinini, Gualtiero (2015) *Physical Computation: A Mechanistic Account*. Clarendon: Oxford University Press.

Piccinini, Gualtiero and Craver, Carl F. (2011) 'Integrating psychology and neuroscience: Functional analyses as mechanism sketches', *Synthese*, 183(3): 283–311.

Piccinini, Gualtiero and Corey Maley. (2014) 'The metaphysics of mind and the multiple sources of multiple realizability', in M. Sprevak and J. Kallestrup (eds.), *New Waves in the Philosophy of Mind*. Palgrave Macmillan. pp. 125–152.

Polger, Thomas W. and Shapiro, Lawrence A. (2008) ‘Understanding the dimensions of realization’, *The Journal of Philosophy*, 105(4): 213-222.

Potochnik, Angela (2016) ‘Scientific explanation: Putting communication first’, *Philosophy of Science*, 83(5): 721-732.

Povich, Mark (2015) 'Mechanisms and model-based functional magnetic resonance imaging', *Philosophy of Science*, 82(5): 1035–46.

Povich, Mark (2018) ‘Minimal models and the generalized ontic conception of scientific explanation’, *The British Journal for the Philosophy of Science*, 69(1), 117-137.

Povich, Mark (2019) ‘Model-based cognitive neuroscience: Multifield mechanistic integration in practice’, *Theory & Psychology*. https://doi.org/10.1177/0959354319863880

Povich, Mark (forthcoming) ‘The narrow ontic counterfactual account of distinctively mathematical explanation’, *The British Journal for the Philosophy of Science.* https://doi.org/10.1093/bjps/axz008

Povich, Mark and Craver, Carl F. (2017) 'Mechanistic levels, reduction, and emergence', in Stuart Glennan and Phyllis Illari (eds.), *The Routledge Handbook of Mechanisms and Mechanical Philosophy*, Routledge. (pp. 185-197).

Prychitko, Emily (2021) ‘The causal situationist account of constitutive relevance’, *Synthese*, 198: 1829–1843

Rice, Collin (2015) 'Moving beyond causes: Optimality models and scientific explanation', *Noûs*, 49(3): 589–615.

Romero, Felipe (2015) 'Why there isn’t inter-level causation in mechanisms', *Synthese*, 192(11): 3731–55.

Ross, Lauren N. (2015) 'Dynamical models and explanation in neuroscience', *Philosophy of Science*, 81(1): 32–54.

Salmon, Wesley C. (1984) *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Salmon, Wesley C. (1989) 'Four decades of scientific explanation', in Wesley C. Salmon and Philip Kitcher (eds.), *Minnesota Studies in the Philosophy of Science, Vol 13: Scientific Explanation*. Minneapolis: University of Minnesota Press. pp. 3–219.

Schmidt, R.C., Carello, Claudia, and Turvey, M.T. (1990) 'Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people', *Journal of Experimental Psychology: Human Perception and Performance*, (16)2: 227–47.

Scriven, Michael (1959) 'Explanation and prediction in evolutionary theory', *Science*, (130)3374: 477–82.

Scriven, Michael (1975) 'Causation as explanation', *Noûs*, 9(1): 3–16.

Shoemaker, Sydney (2007) *Physical Realization*. Oxford: Oxford University Press.

Skow, Bradford (2014) 'Are there non-causal explanations (of particular events)?', *The British Journal for the Philosophy of Science*, 65(3): 445–67.

Smith, Linda B. and Thelen, Esther (2003) ‘Development as a dynamic system’, *Trends in Cognitive Sciences*, 7(8): 343–348.

Stepp, Nigel, Chemero, Anthony, and Turvey, Michael (2011) 'Philosophy for the rest of cognitive science', *Topics in Cognitive Science*, 3(2): 425–37.

Strevens, Michael (2008) *Depth: An Account of Scientific Explanation*. Cambridge: Harvard University Press.

Thelen, Esther, Schöner, Gregor, Scheier, Christian, and Smith, Linda B. (2001) ‘The dynamics of embodiment: A field theory of infant perseverative reaching’, *Behavioral and Brain Sciences*, 24: 1–34.

van Eck, Dingmar (2018) ‘Rethinking the explanatory power of dynamical models in cognitive science’, *Philosophical Psychology*, 31(8): 1131-1161.

van Gelder, Timothy (1998) 'The dynamical hypothesis in cognitive science', *Behavioral and Brain Sciences*, 21(5): 615–28.

van Gelder, Timothy and Port, Robert F. (1995) 'It's about time: An overview of the dynamical approach to cognition', in Robert F. Port and Timothy van Gelder (eds.), *Mind as Motion: Explorations in the Dynamics of Cognition*. Cambridge, MA: The MIT Press. pp 1–43.

Voss, Henning U. (2000) ‘Anticipating chaotic synchronization’, *Physical Review E*, 61: 5115–5519.

Walmsley, Joel (2008) 'Explanation in dynamical cognitive science', *Minds and Machines*, 18(3): 331–48.

Weisberg, Michael (2013) *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

Weiskopf, Daniel A. (2011) 'Models and mechanisms in psychological explanation', *Synthese*, 183(3): 313–38.

Weiskopf, Daniel A. (2017) 'The explanatory autonomy of cognitive models,' in David M. Kaplan (ed.), *Explanation and Integration in Mind and Brain Science*. Oxford: Oxford University Press, (pp. 44-69).

Woodward, James (2003) *Making Things Happen*. Oxford: Oxford University Press.

Woodward, James (2013) 'Mechanistic explanation: Its scope and limits', *Proceedings of the Aristotelian Society Supplementary Volume*, 87(1): 39–65.

Wright, Cory D. (2012) 'Mechanistic explanation without the ontic conception', *European Journal of*

Philosophy of Science, 2(3): 375–94.

Ylikoski, Petri and Kuorikoski, Jaakko (2010) 'Dissecting explanatory power', *Philosophical Studies*, 148(2): 201–19.

Zednik, Carlos (2011) 'The nature of dynamical explanation', *Philosophy of Science*, 78(2): 238–63.

Zednik, Carlos (2015) 'Heuristics, descriptions, and the scope of mechanistic explanation', in C. Malaterre and P.-A. Braillard (Eds.), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences.* Dordrecht: Springer. pp. 295–318.