A field guide to mechanisms: part II

Holly Andersen
Simon Fraser University
handerse@sfu.ca

Abstract: In this field guide, I distinguish five separate senses with which the term 'mechanism' is used in contemporary philosophy of science. Many of these senses have overlapping areas of application but involve distinct philosophical claims and characterize the target mechanisms in relevantly different ways. This field guide will clarify the key features of each sense and introduce some main debates, distinguishing those that transpire within a given sense from those that are best understood as concerning two distinct senses. The 'new mechanisms' sense is the primary sense from which other senses will be distinguished. In part II of this field guide, I consider three further senses of the term that are ontologically 'flat', or at least not explicitly hierarchical in character: equations in structural equation models of causation; causal-physical processes; and information-theoretic constraints on states available to systems. After characterizing each sense, I clarify its ontological commitments, its methodological implications, how it figures in explanations, its implications for reduction, and the key manners in which it differs from other senses of mechanism. I conclude that there is no substantive core meaning shared by all senses, and that debates in contemporary philosophy of science can benefit from clarification regarding precisely which sense of mechanism is at stake.

Keywords: mechanisms; causation; explanation; reduction; methodology; causal processes; interventions; information

0) Introduction

Both mechanisms and causation have become central topics in philosophy of science in recent decades. These two topics are closely connected, and both involved in an intricate nexus of issues concerning explanation, methodology, metaphysics, and more. This field guide addresses five distinct senses in which the term 'mechanism' is discussed in contemporary philosophy of science. The first two senses, covered in part I, are both anti-reductive and explicitly layered in character. In this part II, I consider three additional senses that are 'flat' in comparison: they do not involve mechanistic relationships between levels, and could be treated as compatible with some form of reductionism. Each of these three can be easily confused with the

dominant sense of 'new mechanisms' that has been discussed at length in philosophy of science. My goal here is not to fully develop each of these three senses, but to mark out the main characteristics by which to distinguish them from the first sense in particular (see part I).

Mechanism$_1$, as a reminder, focuses on mechanisms in the sciences. Entities and activities are organized such that specific conditions trigger interactions between entities that progress through a regular series of activities and either causally produce or give rise to by constituting the phenomenon to be explained. Mechanisms$_1$ nest, such that subsections of a mechanism will themselves be constituted by some lower-level mechanism, without thereby reducing to that mechanism. Mechanism$_2$ is the ontological claim that the world is fully mechanistic in character, comprised of hierarchical causal layers that are constituted out of mechanisms at various levels. Mechanism$_2$ differ from mechanism$_1$ largely in that the latter, but not the former, involves a requirement that mechanisms occur with some form of regularity. Mechanism$_2$ bears a much closer relationship to the historical discussion of mechanism, but has far fewer direct methodological consequences for the sciences, than does mechanism$_1$.

The main contrast between mechanism$_{1or2}$ and the three senses to be presented here, is whether the senses explicitly acknowledge a certain kind of tiered structure to the world. The first two senses both address sub-mechanisms nested within mechanisms, which allows for mechanisms to individuate levels in the world and for connections to be drawn between mechanisms at different levels without thereby committing to the reduction of higher levels to lower levels. As such, mechanism$_1$ and mechanism$_2$ open up the possibility for meaningful ontological significance of higher-level entities and activities. In contrast, the next three senses do not explicitly address levels in the world: they are ontologically flat. This does not mean that any of senses three through five are thereby reductive in character. They may allow for an ontology containing higher-level entities or activities, such as mechanism$_4$. They may utilize higher-level causal relata, such as mechanism$_3$, but

without explicitly connecting them to lower level relata. As such, each of these three senses is compatible with either a reductive or anti-reductive ontology. This is especially relevant to debates about higher level causation, since two of the major senses to be discussed in this paper are elements in major contemporary accounts of causation.

1) Mechanism$_3$: Mechanisms as single equations in structural equation models

This species of mechanism$_3$ can be found within structural equation approaches to modeling causation. Such accounts characterize a system in terms of a number of variables, both exogenous (not caused by other variables in the system) and endogenous (caused by some other variable in the system). Causal relationships connect variable nodes, and the entire system's causal structure can be given in diagrammatic form in terms of directed arrows between such variable nodes. This graphical representation can be made more specific by adding weights to each of the arrows; this results in a set of equations that are collectively called the structural equation model for that system. A system of variables can be represented with a series of structural equations that specify the value of each variable in terms of the values of its causal ancestors. A mechanism$_3$ is a causal relationship in the world that is represented with a single equation in a set of structural equations representing the system in question. A mechanism$_3$ is defined by its capacity to be manipulated independently of the other causal relationships in the system, in terms of difference-making. This approach presumes that each structural equation corresponds to a mechanism$_3$ that can be intervened on independently of other mechanisms$_3$ in the system.

A major example is Woodward's (2003) interventionist counterfactual account of causation.[1] He defines causation in terms of possible interventions on variables in systems, such that an intervention breaks the existing causal structure to set the value for the intervened-on variable independently of its causal ancestors in the system. Under an appropriate manipulation, further variable value changes can be attributed to the intervention, providing information about the causal effects of the intervened-on variable. Other accounts of structural equation modeling (Spirtes, Glymour, and Scheines 2000; Pearl 2000) posit that causal relationships between variables are revealed by characteristic changes in probabilistic relationships between those variables under intervention.

Woodward (2003) defines modularity as a feature that representations of causal relationships can have, where that feature reflects the independence of causal relationships in the world: "…a system of equations will be modular if it is possible to disrupt or replace (the relationships represented by) any one of the equations in the system by means of an intervention on (the magnitude corresponding to) the dependent variable in that equation, without disrupting any of the other equations" (48). Modularity of equations is taken as an indicator that the model has accurately isolated robust causal relationships that can be used to change effect variables by intervening in cause variables without destroying the relationship between cause and effect.[2] When we have a set of modular structural equations, they each represent a different, independent, causal relationship that is called a mechanism$_3$.

Mechanism$_3$ crucially involves the assumption that a full and correct representation of a causal system will be modular. Justification for this assumption is based on the nature of the underlying causal structure represented by the modular equations,

[1] It is worth noting that while Woodward (2003) provides a clear statement of mechanism$_3$, Woodward has also written elsewhere about mechanisms in ways that count as mechanism$_1$ (e.g. Woodward 2002).
[2] This is Woodward's characterization of modularity, since his account is the primary illustration in this section. There are additional related but distinct characterizations of modularity in the literature.

namely, on consideration of a mechanism$_3$. "In what follows, I assume that when causal relationships are correctly and fully represented by systems of equations, each equation will correspond to a distinct causal mechanism and that the equation system will be modular" (Woodward 2003, 49). In other words, causal mechanisms$_3$ are independent of one another when accurately picked out, and so an adequate representation of them will involve modular structural equations. This view of mechanisms is compatible with a range of mutually incompatible ontological views, but does offer two basic constraints on ontology: one, that mechanisms$_3$ be independently manipulable; and two, that causation is difference-making or counterfactual in character (as opposed to, for instance, productive or actual – see mechanism$_4$).

Compared with its weak ontological commitments, this view of mechanism comes as part of a rich methodological package. Indeed, it is defined by its place in that methodological package. Distinguishing interventions from observations, plus drawing on the carefully developed mathematical techniques embodied in structural equation models, have been a huge breakthrough for the discovery of causal structure from data. Explanatorily, it is part of a specific view of causal explanations: that they involve counterfactuals, that they have specific and empirical consequences that can be confirmed or disconfirmed, and so on. This should not be confused with the view that all explanations are causal explanations. Rather, it is the view that, regardless of what noncausal explanations look like, causal explanations take a certain form. Finally, it is agnostic with respect to reduction. The models for individual systems tend to be ontologically 'flat', where the variables are all at similar levels (of size, organization, or other level differentiation). One can use this model with the added assumption that the causal relationships represented by it are genuine; by doing so, one is attributing genuine causation to higher-level relata, thus committing to an anti-reductive view about causation. Or, one could use the very same model, in the same ways, with a commitment to reduction and/or microphysicalism about causation: in that case, one is committed to the view that

the represented relationships are not the 'real' causal story, but that they are merely a useful stand-in.

The key features that indicate a mechanism$_3$, modularity of mathematical representation and manipulability independent from other mechanisms$_3$ in the system, are relevantly different selection criteria than for mechanism$_1$ or mechanism$_2$. It is possible to meet the criteria to be a mechanism$_1$ without thereby meeting the independence requirements to be a mechanism$_3$; it is also possible for a system to meet the criteria to be a mechanism$_3$ without thereby meeting the criteria to be a mechanism$_1$. Two interesting examples illustrate these claims: gene knock-out experiments, and the Hodgkin and Huxley model of action potentials in neurons. These examples also demonstrate how clarifying the difference between distinct senses of mechanism can benefit certain debates that turn out to involve conflicting claims about what counts as a mechanism.

The first example shows how a mechanism$_1$ can fail to be a mechanism$_3$. Mitchell (2008) takes the case of gene knock-out experiments to show how certain commitments of the interventionist account are undermined by experiments on complex systems. Her argument is that the causal relationships in these systems are not independent in the requisite way and thus cannot be represented with modular equations, regardless of how one rearranges the representation. Intervening on one causal relationship actually changes the structure of other relationships, such that none can be intervened on with the requisite independence and the system cannot have an accurate and modular representation. Mechanisms$_1$ are likely involved in this system-wide rearrangement of causal structure, but the system fails to have mechanisms$_3$.[3]

Mitchell (2008) considers genetic engineering as an intervention to investigate causal relationships between specific genes and phenotype, by removing a gene

[3] This labeling of mechanism types, it should be understood, is my own; Mitchell does not specifically address this point.

from the genome to see what phenotypic effects are observed. The problem is that in up to 30% of gene knockout experiments, no effect is observed on phenotype. This appears to be due to the plasticity and robustness of the genome in reorganizing, using redundancy and degeneracy, in order to preserve phenotype under alteration of formerly causally relevant genes.

> Redundancy in a genetic regulatory system describes a situation where there are multiple copies of a functional gene. Redundancy is widespread in biological systems, though it is somewhat puzzling how it could have evolved. … Degeneracy, as defined by Edelman and Gally (2001), is distinguished from copy redundancy. It refers to an organization where alternative components and structures with distinct functions may nevertheless produce the effect of a targeted component when the component is no longer operative. Degeneracy has been identified in 22 different levels of biological systems… (2008, 701)

The upshot of this robust reorganizational ability is that in degenerate systems, distinct causal pathways cannot be disrupted independently. Consequently, they cannot be represented as modular, no matter how accurate or fine-grained our representation. Nonetheless, these cases meet the criteria for a mechanism$_1$ as outlined in part I of this field guide: they are comprised of entities engaged in very complex activities, organized in such ways that they reliably produce certain regularities once triggering conditions occur. These mechanisms$_1$ are robust enough to produce these regularities in phenotype under wide variation in start-up conditions and internal organization of entities, by compensating in terms of new activities by remaining entities.

The next example shows how a mechanism$_3$ may fail to be a mechanism$_1$. Weber (2008) considers the Hodgkin and Huxley model of action potentials in neurons. This model is a set of equations that describe the time course of the action potential in terms of currents from voltage-gated ion channels. The equations were developed to match the time-indexed data from giant squid axons. Because the equations were

designed to match a specific set of empirical observations, however, generalizing them is difficult. There is little to no reason to think that these equations correspond to any mechanism$_1$ in axons. Rather, they most likely aggregate the action of a number, potentially a very large number, of mechanisms$_1$ involved in action potential propagation. The potentially uncoordinated collective results of many different mechanisms$_1$ are captured by the equations, with the consequence that changes in any number of mechanisms$_1$ result in unpredictable changes to the outcome.

Bogen (2005) argues that the Hodgkin and Huxley model cannot provide a genuine mechanism$_1$: because the weighting constants in the equations were calculated solely to match the data, it is a mathematical convenience instead of evidence of a real mechanism. Craver (2006) [4] argues that the H-H model is (at most) a mechanism$_1$ sketch, providing a constraint on the space of possible mechanisms that could give rise to the regularity described by the equations. While the H-H model describes a regularity to be explained with mechanisms$_1$, he says, it is insufficient to pick out such a mechanism uniquely. We still lack knowledge of the component entities, the activities in which they are engaging, or the organization required for the action potential to take the form described by the equations. If the H-H model represents the summed actions of multiple distinct mechanisms$_1$, there may not be an overarching coordination between those mechanisms. The physical possibilities compatible with the mathematical equations are too broad.

Weber (2006) agrees that the H-H model does not count as a mechanism "in the narrow sense recently discussed in the philosophy of biology and neuroscience" (2), namely, as a mechanism$_1$. However, he argues that it does meet the criteria to count as genuinely causal according to the criteria for a mechanism$_3$ in Woodward (2003). The H-H model is modular, and also invariant under many interventions. Weber

[4] Again, it should be noted that Bogen's and Craver's argument are in terms of mechanism simpliciter; I have construed their arguments in the terminology of this field guide as part of illustrating the difference between these two senses.

concludes: “The HH equations thus satisfy all the conditions that a major recent theory of causation and explanation requires from causal-explanatory generalizations. There is simply no conflict between a generalization being causal and explanatory and it's being an experimental generalization or being fitted to data” (2008, 1003). Weber’s argument does not address the lack of a mechanism$_1$, but rather points to the usefulness of a mechanism$_3$ for explanation.

Distinguishing between mechanism$_1$ and mechanism$_3$ is helpful to these debates in philosophy of science, and generally for debates in which mechanisms figure in causal or explanatory claims. Even though both play important roles in science and in giving causal explanations, mechanism$_1$ and mechanism$_3$ pick out different structures in the world, and involve different methodological considerations for how to investigate those mechanisms.

Finally, note how mechanism$_3$ is a clear and viable use of the term mechanism, occurring within an account of causation, where that account of causation is not itself mechanistic. This marks a strong distinction between mechanism$_2$ and mechanism$_3$: the former is a mechanistic account of causation, the latter is part of a counterfactual account of causation. Woodward’s influential (2003) interventionist account of causation defines causal claims at least partially in terms of counterfactuals; neither a mechanism$_1$ nor a mechanism$_2$ are required. Mechanism$_2$ is part of a worldview where causation is physical and productive, whereas mechanism$_3$ is part of a difference-making theory of causation. Mechanism$_2$ and mechanism$_3$ must be clearly distinguished because they are committed to different methodologies for finding causal relationships, to different evidentiary standards for establishing the truth or falsity of causal claims, and to basic differences in the ontology of causation.

2) Mechanism$_4$: Mechanisms as processes and/or interactions in physical process accounts of causation

The term 'causal mechanism' is sometimes used in process accounts of causation. Mechanisms$_4$ are the space-time processes that can have causal effects on or be causally affected by other causal processes/mechanisms$_4$. They are often, although not necessarily, taken to be microphysical in character, and their interactions are often construed in terms taken from physics.

Mechanisms$_4$ are the main elements in one of the major physical, mechanistic, accounts of causation, strongly associated with Salmon (1998, 1994, 1977) and Dowe (2000). Salmon offered a theory of causation in which causal processes propagate through time and space by being at points in space at successive moments in time. Causal processes are distinguished by their ability to bear marks with them through time and interact with one another; pseudo-processes cannot bear marks through time, and cannot engage in genuine causal interactions. Causal interactions occur when two or more causal processes intersect in space and time and exchange marks. "Causal processes and causal interactions are the basic *causal mechanisms* according to this approach" (Salmon 1994, 237; italics added). Dowe (2000) updated Salmon's account by relying on the criterion of transference of conserved physical quantities instead of mark transmission. Causal processes are those capable of transmitting conserved quantities such as energy or charge; causal interactions become exchanges of conserved quantities between two or more causal processes. Both causal processes and interactions, taken together, are mechanisms$_4$.

This account of physical causation was developed specifically as an alternative to accounts of causation involving counterfactuals (see especially Salmon 1994). Physical causation has truth conditions that are clearly and unproblematically actualized in this world. We can trace out causal relationships in the world by tracking causal mechanisms$_4$, namely, by tracking conserved quantities through causal interactions. Evaluation of the truth of causal claims thus avoids problematic metaphysical baggage like possible worlds. Thus, the ontological commitments of this view are firm and part of the attraction it holds for some philosophers.

Methodologically, mechanism$_4$ offers extremely specific guidance for investigating mechanisms under some conditions, and no guidance whatsoever under other conditions. In Dowe's version, mechanisms$_4$ can be traced out exhaustively by keeping track through time and space of the quantities that modern physics tells us are conserved. When it is possible to do this, the methodological implications could not be clearer. In contrast, for causal relationships that do not apparently involve physical processes that can be spatiotemporally tracked, there are no methodological implications. If we are considering possible causal relationships between, for instance, natural resources, democratic governance, and poverty, this view has very little to offer. According to mechanism$_4$, apparent explanations involving relata such as poverty and democratic governance are not genuinely causal, and thus are not genuinely explanatory. They are at most stand-ins for some more complicated microphysical story, the actual details of which are left entirely open.

The issues with mechanisms$_4$ and reduction are carefully treated in Williamson (2011). He points out two main problems with process theories of causation: the mechanisms in question, involving conserved quantities, may be too 'large' to accommodate much of physics, including quantum mechanics; and they may also be too 'small' to accommodate causal relationships in higher-level sciences. A consequence of Williamson's criticism is that treating these mechanisms as constitutive of causation undermines any potential causal autonomy of higher-level causes posited by the special sciences. Even though it is not explicitly reductive, it may nevertheless have reductive consequences.

This brings us to the importance of distinguishing mechanism$_4$ and mechanism$_{1or2}$. Mechanism$_4$ is notoriously unable to accommodate disconnection, where a spatiotemporal gap in what we ordinarily consider to be a legitimate causal relationship means that there is no mechanism$_4$. Both mechanism$_1$ and mechanism$_2$ can easily incorporate disconnections as part of the activities or of the organization

in the mechanism. Mechanism$_3$ and mechanism$_4$ are both accounts of causation, but have deeply incompatible ontological commitments. Salmon offered mechanism$_4$ as an explicit alternative to counterfactuals. Woodward's account of causation is counterfactual and not mechanistic; he explicitly denies Salmon's claim that spatio-temporal continuity is required for causation (Woodward 2003, 147).

Even though mechanism$_2$ and mechanism$_4$ are both broadly mechanistic accounts of causation, they differ sharply in their ontological commitments about the structure of the world and its fundamental units. One of the strengths of mechanism$_2$ is its ability to flexibly pick out appropriate physical grain size for various mechanisms, and to nest those mechanisms hierarchically as both partially constituting higher-level mechanisms and as constituted by lower-level mechanisms. Mechanism$_4$ lacks this feature (Williamson 2011). This means that while mechanism$_2$ effectively blocks reduction of higher-level causal relata, mechanism$_4$ points towards microphysicalism about causation.

3) Mechanism$_5$: Mechanisms as net constraints on the states open to a system

Mechanism$_5$ is not currently a widely used sense of the term, but as more philosophers of science and mind become aware of the integrated information account of consciousness (Tononi 2004, 2009; Balduzzi and Tononi 2008, 2009; Tononi and Koch 2008), it will become more relevant. A mechanism$_5$ is defined in an information-theoretic system, and is the basis for distinguishing probability distributions over states in that system; differences in these distributions, the result of the mechanism$_5$, are used to calculate various information-theoretic relationships. A mechanism$_5$ just is the net constraint that individual node values in a system place on connected nodes. It is a mathematical characterization; there can be mechanisms$_5$ in purely abstract systems that represent nothing physical. This sense of mechanism stands apart from the other senses, insofar as there is nothing causal involved in mechanism$_5$. Mechanism$_5$ is not something that could be given as

an explanation for a phenomenon – it is instead a precise description of a phenomenon that requires explanation.

Tononi's characterization of a mechanism$_5$ is a short part of a dense and highly mathematized account of consciousness and phenomenal qualities. The mechanism$_5$, as Tononi describes it (e.g. Balduzzi and Tononi 2009,Tononi 2009), is the set of constraints of node values on node values that determine the actual repertoire of states instantiated in the brain. In integrated information theory of consciousness, that makes the mechanism$_5$ the pattern by which states in the potential repertoire are eliminated as possible states the brain can enter, based on its current state.

> Elements are linked by connections to form a directed graph, specifying which source elements are capable of affecting which target elements. Each target element is endowed with a ''mechanism'' or rule through which it determines its next output based on the inputs it receives. These mechanisms are assumed to be elementary, for example AND, XOR; they can also be probabilistic. … No perturbation can be ruled out a priori, since it is only by passing a state through the mechanism that the system generates information. (Tononi 2009, 3)

These constraints can shift over time: Tononi assumes that there are ongoing changes to the constraints by one node state on neighboring node states. Node 1 in state 1 at time 1 may constrain node 2 to only one possible value, but then node 1 in state 1 at time 10 may not constrain node 2 values at all.

To see why mechanism$_5$ is acausal, it is helpful to see how it is derived. Consider a connected set of nodes, where the nodes can take different values that represent different states of the system at that point, and connections represent an abstract relationship between the nodes (these should not be simply treated as causal relationships – causal relationships are a proper subset of the relationships that

might connect nodes in the information-theoretic sense). The potential repertoire of a system is the total number of distinct states in which the system could be found (see Cover and Thomas 2006, Applebaum 2008). This is simply a factorial of all the different states each node could be in, taken over all of the nodes in the system. It is the largest possible number of states for the system, treating all states as accessible to the system (and, usually, treating all states as equally probable). For a complex network with many nodes, this is a very high number.

The actual repertoire of states open to the system, however, may be smaller than its potential repertoire. This occurs when there are internal constraints on the states that the system may enter into based on the state it is currently in. The actual repertoire reflects the ways in which node values at one time constrain the values of neighboring nodes at later times. Both the potential and actual repertoires can be represented with a probability distribution over node values. When the difference between the potential and actual repertoires is small (considered in orders of magnitude), the relative entropy between the two distributions is low, which means that the information contained in the system's state at any given moment is also low. When the actual repertoire is quite small relative to the potential repertoire, such that node values highly constrain other node values, then the relative entropy of the two probability distributions is high, and actual state of the system contains a great deal of information.

Mechanism$_5$ just is the description of those constraints on node values; it does not presume anything about a physical or causal reason for those constraints being what they are. Mechanism$_5$ is a precise characterization of the repertoire difference that allows a system to bear information. It cannot explain that pattern, but is instead the pattern to be explained.

There are, oddly enough, no ontological commitments associated with this sense of mechanism. A mechanism$_5$ is defined as whatever the current pattern of constraint is that yields the actual repertoire of the system at a given moment. This is a

mathematical characterization of mechanisms; there need be no physical system at all. One can simply define a set of constraint parameters on an abstract system of nodes and have a mechanism$_5$. Even when a physical system is being represented, such as the brain, there is no reason to think the mechanism$_5$ maps onto or results from physically significant features of the system. They may simply be the cumulative result of noise in the system at that time, reflecting nothing of underlying structure or causal relationships.[5]

Methodologically, information theory provides powerful ways to find a variety of subtle mathematical relationships within defined systems. How this extends to the physical world will depend on details about the system being represented and how veridically the mathematical model represents the system in question. Explanatorily, mechanism$_5$ is opaque: its definitional character means that providing the mechanism$_5$ for a system does not explain anything about the behavior of the system. It describes, in a different format, that which requires explanation.

The issue of reduction is best addressed by clarifying the acausal character of mechanism$_5$. It renders this sense radically differs from the others, especially from mechanism$_{1or2}$. All the other senses of mechanism can be offered as explanations for phenomena, including phenomena that are higher-level than the mechanism in question; this explanatory character is generally taken as a hallmark feature of mechanisms. In contrast, mechanism$_5$ is the explanandum, requiring such an explanation. Our initial tendency may be to assume that there is some underlying mechanism that *results in* the constraint pattern that gives rise to the actual repertoire. This would be a misunderstanding. Tononi himself may perpetuate this misunderstanding, as he occasionally refers to these as causal mechanisms (e.g.

[5] Tononi's use of mechanisms for his account of consciousness is part of a set of ontological commitments about the nature of consciousness. Given the mathematical character of his exposition of the theory, however, with little to no physical detail about how this is implemented in the brain, it is rather unclear what this amounts to.

2009, 3). However, when you consider the role they play in the information-theoretic framework, it becomes clear that while these 'mechanisms' could be supplemented by or given rise to 'mechanisms' in some causal sense, they cannot simply be those mechanisms. Conversely, they need not be instantiated by mechanisms in any of the other senses, but could instead be purely mathematical and formal in character.

There very well may be some mechanism, in some other sense, underlying the constraint pattern in a physical system at any given moment; there is potentially an enormous number of them. However, in this terminology, the pattern of constraints constitutes the mechanism$_5$; it is not the result of the mechanism, not what is explained by or accounted for by the mechanism, and not the end product of the mechanism. This means that mechanism$_5$ is diametrically opposed to mechanisms$_1$ (as well as, for slightly different reasons, mechanisms$_{2\text{-}4}$) in that mechanism$_1$ takes such patterns or regularities as the phenomena to be explained by a mechanism, whereas mechanism$_5$ takes the pattern or regularity itself to be the mechanism.

4) Conclusion

This part II of the field guide to mechanisms has considered three varieties of mechanisms that are not explicitly anti-reductionist in character. Each of the three senses discussed here are somewhat less common in contemporary philosophy of science than the 'new mechanisms' approach (mechanism$_1$). Part of my goal in this paper has been to offer a helpful outline of many discussions of mechanisms in contemporary philosophy of science, since many outside this field are interested in the new developments on this front. Within philosophy of science, I am arguing that care must be taken to explicitly distinguish which sense is at play in a given debate. This means that philosophers of science who want to extend one sense of mechanism to a new context should provide clear justification for why it is in fact

the same notion of mechanism that they are using, rather than a related but different idea with the same name.

In elaborating five distinct senses, I've argued that each sense has a unique set of ontological, methodological, explanatory, and anti-/reductive commitments. There is no single common core that all these senses of mechanism share, such that we might treat this as the unifying meaning of the term. There are no common criteria for what is to count as a mechanism, no common set of constituents that make up a mechanism, no common explanatory role played by all. Instead of looking for a common core meaning, it is illuminating to compare the regards in which senses of mechanism differ or overlap. Some senses of mechanism are closer in kind to one another: for instance, both mechanism$_2$ and mechanism$_4$ offer productive, mechanistic accounts of causation. In contrast, mechanism$_3$ is part of a difference-making account of causation that is anti-mechanistic. Mechanism$_1$ is compatible with senses 2, 3, or 4 as a way of fleshing out the causal activities within mechanisms$_1$. Mechanism$_5$ requires supplementation by some other sense of mechanism in order to explain the behavior of systems it describes. Recognizing the distinctions between these different notions of mechanism will help clarify claims made regarding mechanisms, and will help resolve or dissolve some disagreements that turn on different characterizations all going under the heading of 'mechanism'.

References

Applebaum, David (2008). *Probability and information: An integrated approach*. (2nd ed.). Cambridge: Cambridge University Press.

Balduzzi, David and Tononi, Giulio (2008). "Integrated Information in Discrete Dynamical Systems: Motivation and Theoretical Framework" *PLoS Computational Biology* 4:6, e1000091.

Balduzzi, David and Tononi, Giulio (2009). "Qualia: The Geometry of Integrated Information" *PLoS Computational Biology* 5:68 e1000462.

Bogen, Jim (2005). "Regularities and Causality; Generalizations and Causal Explanations." *Studies in History and Philosophy of Biological and Biomedical Science* 36: 397-420.

Cover, Thomas, and Thomas, Joy (2006). *Elements of Information Theory* 2nd edition. Wiley-Interscience: Hoboken, NJ.

Craver, Carl (2006). "When Mechanistic Models Explain." *Synthese* 153: 355–376.

Dowe, Phil (2000) *Physical Causation*. Cambridge University Press.

Mitchell, Sandra D. (2008). "Exporting Causal Knowledge in Evolutionary and Developmental Biology." *Philosophy of Science* 75(5): 697-706.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Salmon, Wesley (1977). "An At-At Theory of Causal Influence." *Philosophy of Science* 44(2): 215-224.

Salmon, Wesley (1994). "Causality Without Counterfactuals." *Philosophy of Science* 61 (2): 297-312.

Salmon, Wesley (1998). *Causality and Explanation*. Oxford University Press.

Spirtes, Peter, Clark Glymour, and Richard Scheines. 2000. *Causation, Prediction, and Search*. Cambridge, MA: The MIT Press.

Tononi, Giulio (2004). "An information integration theory of consciousness." *BMC Neuroscience* 5 (42): doi:10.1186/1471-2202-5-42.

Tononi, Giulio (2009). "Information Integration Theory" in *The Oxford Companion to Consciousness*, Bayne, Cleeremans, and Wilkins (eds). Oxford: Oxford University Press, 380.

Tononi, Giulio, and Koch, Christof (2008). "The neural correlates of consciousness: an update." *Annals of the New York Academy of Sciences* 1124: 239–261.

Weber, Marcel (2006) *Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation.* In: [2006] Philosophy of Science Assoc. 20th Biennial Mtg (Vancouver) > PSA 2006 Symposia. http://philsci-archive.pitt.edu/3287/

Weber, Marcel (2008) "Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation." *Philosophy of Science* , Vol. 75, No. 5, Proceedings of the 2006 Biennial Meeting of the Philosophy of Science Association Part II: Symposia Papers, eds. Bicchieri and Alexander, 995-1007

Williamson, Jon (2011). "Mechanistic Theories of Causality Part I." *Philosophy Compass* 6 (6): 421-432.

Woodward, James (2002). "What is a Mechanism? A Counterfactual Account." *Philosophy of Science* 69 (S3): S366-S377.

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