Draft Symposium Paper for PSA 2016
Please do not quote without permission

Towards Mechanism 2.0: Expanding the Scope of Mechanistic Explanation

Arnon Levy
Hebrew University of Jerusalem

William Bechtel
University of California, San Diego

1. Introduction

Accounts of mechanistic explanation, especially as applied to biology and sometimes going under the heading of "new mechanism," provided an attractive alternative to nomological accounts that preceded them. These accounts were motivated by selected examples, drawn primarily from cell and molecular biology and neuroscience,. These examples pointed to sharp contrasts between real-life biological explanation and discovery and the then-dominant models of scientific explanation. However, the range of examples that scientists take to be mechanistic explanations is far broader. We focus on examples that differ from those traditionally recruited by Mechanists. Our contention is that attention to additional examples will lead to a richer conception of mechanistic explanation, prompting a shift from what we refer to as Mechanism 1.0 to Mechanism 2.0. In suggesting such a move, our goal is not to downplay the importance of Mechanism 1.0 and the progress it signified. Mechanism was a big step forward in philosophy of biology. We just think it's time to take the next step. Furthermore, by adopting the language of Mechanism 1.0 and 2.0 we mean to signal that we anticipate further enhancements to the conception of mechanistic explanation as philosophers of science attend to more examples of scientists advancing what they characterize as mechanistic explanations.

One way to approach the distinction between Mechanism 1.0 and 2.0 is to return to the machine metaphor that inspired mechanistic research in biology and was invoked explicitly by Bechtel and Richardson (1993/2010) and implicitly - primarily in the choice of examples - by other writers on mechanistic explanation. Most mechanists have attempted to differentiate biological mechanisms from machines. However, the examples used to motivate accounts of mechanistic explanation by, for example, Bechtel & Richardson (1993/2010), Bechtel (2006), Machamer, Darden, and Craver (2000), and Craver and Darden (2013) are in fact much like traditional machines. Thus, protein synthesis is presented as involving the creation of mRNA from the DNA template in the nucleus and the transport of the mRNA to the ribosome, where it serves as a template for forming a chain of amino acids. Oxidative metabolism is described as localized to the mitochondria of cells where a specific set of enzymes catalyze the successive oxidation of metabolites until only carbon dioxide and water remain, generating ATP in the process. In these examples, the

mechanism consists of a bounded set of enduring entities or parts in a fixed configuration. These explanations accord central importance to the structure of parts and often envision the mechanism's organization as sequential, or perhaps branching. As in classical machines—steam engines, typewriters and food processors—the parts are envisaged as performing the same activities or operations every time they are called upon so as to produce the phenomenon to be explained.

These features of the examples, we contend, did much to advance an attractive and compelling picture of explanation that attracted much interest. For example, they established that such explanations did not fit the D-N model. By portraying a sequence of operations that resulted in storing energy in ATP or synthesizing proteins, scientists explained these processes without explicitly invoking laws. Moreover, for philosophers who were expanding their focus beyond justification to include discovery, these examples provided case studies of how the two practices were connected (Darden, 2006). The account also offers norms of success: to understand how a system in nature works, one should be able to identify its parts, demonstrate what operations they perform, and describe how they are organized so as to work together. If researchers cannot identify parts and trace how they operate on each other, they fail to show how the mechanism generated the phenomenon in terms of parts.

Still working within the framework of Mechanism 1.0, some philosophers began to focus on biological mechanisms whose parts are organized in a more complex manner (e.g., into multiple feedback loops). This undercut the ability of scientists to mentally rehearse the functioning of the mechanism to understand how it brought about its behavior. Instead, scientists had to appeal to mathematical representations and perform computational simulations. When the required mathematical representations are non-linear, computational simulations show how mechanisms can produce complex behavior (e.g., oscillations that partly synchronize with each other). While mechanists such as Bechtel and Abrahamsen (2010) and Brigandt (2013) distinguished such explanations as dynamical mechanistic explanations, they did not fundamentally challenge the framework of mechanism 1.0: they still appealed to a stable and bounded set of parts whose structure determines their interactions. Rather, they took a relatively small step away from Mechanism 1.0 to version 1.1.

Our project is much the same as the philosophers who advanced Mechanism 1.0. We focus on examples of explanation in biology. The difference is that we focus on ones that do not fit the picture of Mechanism 1.0 or 1.1. Some philosophers might see these departures from Mechanism 1.0 as requiring abandoning mechanism altogether. We certainly think that there are explanations that are not mechanistic – such as teleological, etiological and perhaps mathematical explanations. But the cases we explore here, while differing from Mechanism 1.0 in specific respects that we will highlight, are still recognizably parts-and-operations explanations, and are typically characterized as such by scientists. We will identify several ways in which these examples reveal limits of Mechanism 1.0 in the next section. In our view, these examples motivate expanding and reconceptualizing what a

mechanism is (hence, we speak of Mechanism 2.0). Although we are not yet at the point where it makes sense to offer a full characterization of Mechanism 2.0, in section 3 we will both articulate how the examples offered for Mechanism 2.0 enrich our understanding of mechanistic explanation and how mechanisms may differ from machines as they have been traditionally understood. We will also identify work that remains to be done in developing the conception of Mechanism 2.0.

2. The limits of Mechanism 1.0

In this section we will discuss departures from five key aspects of Mechanism 1.0: the appeal to the structural features of parts; the idea of a stable and straightforward organization; the assumption that parts are stable; the idea that mechanisms have well-defined boundaries in space and time; and the conception of mechanisms themselves as enduring entities.

2.1. Parts that are not discrete entities

In the examples used to illustrate and motivate Mechanism 1.0, mechanisms consisted of discrete entities that could be identified structurally in terms of properties such as shape, size, and mass. Indeed, many mechanistic explanations are presented in terms of individual entities such as molecules that undergo transformations (perhaps binding to another molecule and changing their shape in the process). (This is how mechanisms are represented in many diagrams of mechanisms, including those below.) But in fact the working part is very often not a single entity, but a large collection of similarly structured entities. And it is very often not only—sometimes not primarily—structural properties that matters but aggregative features such as the concentrations of these entities that performs the work.

Phenomena that involve electrical potentials over membranes provide examples whose explanations depend on concentrations. For instance, ATP synthesis strongly depends on the direction of the proton gradient across the mitochondrial membrane. Similarly for action potentials: to understand how an action potential works, you can't focus merely on the molecules that are involved (sodium, potassium, ion channels etc.) and their structures. It is the relative concentrations of ions, inside and outside the cell, that determine the timing and size of a spike. In such examples, components with the relevant structures are present throughout the process, on both sides of the membrane. What drives the process isn't their mere presence or structure. It is their relative concentration, which changes continuously as the mechanism operates. As protons build up in the mitochondria's inter-membranal space, they generate a potential, then is then converted into ATP via the ATP synthase "windmill." Structural aspects matter here, of course, but without careful attention to concentrations, the system cannot be understood. The role of concentration is perhaps even more subtle in action potentials. Sodium and potassium build up to a steady state concentration, maintaining a resting potential. When an above-threshold excitation arrives, sodium and potassium change concentrations quickly and in opposite "directions." It is the precise shifts in concentration and their timing that determines whether a spike is

generated. The structures of ions and ions channels, matter, but changes in concentrations are key.[1]

A major focus of contemporary biology is the regulation of biological processes, and here one often finds references to switches. A recent paper by Nathan (2014) looks at such a case in depth, arguing for a notion of causation by concentration. Nathan's analysis is compelling, though he does not explore potential ramifications for Mechanism. We focus on this aspect. A genetic switch is a bi-stable system in which a given gene can be either "on", leading to high levels of transcription, or "off", leading to low levels of transcription. The lac operon and the viral lambda switch are very well studied examples. In genetic switches, it is not merely the ability of an inducer or an inhibiting molecule to bind to DNA, and initiate transcription, that explains switching. Indeed, inducers and inhibitors are typically bound to DNA, to some extent, at all times. But what determines whether a switching event occurs is the relative rate of binding, which is determined by the their relative concentrations. Again, the structure of the parts remains constant. It is the shift in concentrations that moves the system into wholly new states. (For an overview of the principles underlying genetic switching, see Nelson, 2015, Chapter 10.)

The cases presented in this section illustrate that in some explanations what is doing the explanatory work is not the structure of the parts in question, but the concentrations of the parts. Ironically, both ATP synthesis and the generation of action potentials have been used in support of Mechanism 1.0. The fact that in these examples it was the concentration of protons or electrons that mattered was not noted. But it has direct consequences in terms of how the explanation in question is to be confirmed, what kind of discovery strategies it will be linked to and so on. Thus, a move away from the idea of parts as (merely) structural units has important consequences.

2.2. Organization that is not fixed

Although it was typically not commented on, those offering examples of Mechanism 1.0 recognized that, like parts of machines, the parts of a mechanism go through a sequence of different states as the mechanism generates a phenomenon. In many cases, including enzyme-catalyzed metabolic reactions, cell-to-cell signaling, DNA and RNA processing, and various other phenomena, the ability of a part to perform an operation depends on its three-dimensional shape and associated physical features such as its electrical charge. It is as a result of such features that an enzyme is able to bind to a substrate. When bound to a substrate, it is no longer able to bind to another and the conformation of the molecule is changed. Once the reaction is complete, it returns to its initial conformation and is able to bind another molecule of substrate.

[1] This is illustrated by the fact that Hodgkin and Huxley, who knew little about the structures of underlying molecules, still managed to produce a seminal advance in our understanding of action potentials via a model that took into account facts about concentrations alone. See Levy (2013).

[2] In the spirit of our discussion of concentrations above, registration of time of day appears to be a population

While recognizing these local changes of parts and how they interact with each other in the course of a mechanism's operations, Mechanism 1.0 presented the overall organization of mechanisms as unchanging. This is reflected in what is perhaps the standard form in which many mechanisms are represented: a flow diagram in which nodes represent parts and arrows represent operations performed by one part on other parts. Most diagrams do not represent how individual parts change structurally over time. For example, Figure 1, showing the activities of transcription and translation leading to the synthesis of new proteins, only shows tRNAs binding to amino acids and transporting them to the ribosome, not any change in the tRNA that results. Rather, the focus is on tracing how each part interacts with other parts. The various activities that are shown--the assembly of the mRNA along the DNA template, the transport of the mRNA (and tRNA and rRNA) to the cytoplasm, the ferrying of amino acids to mRNAs by tRNAs, and the binding of amino acids are into a polypeptide chain--are presented as enduring.

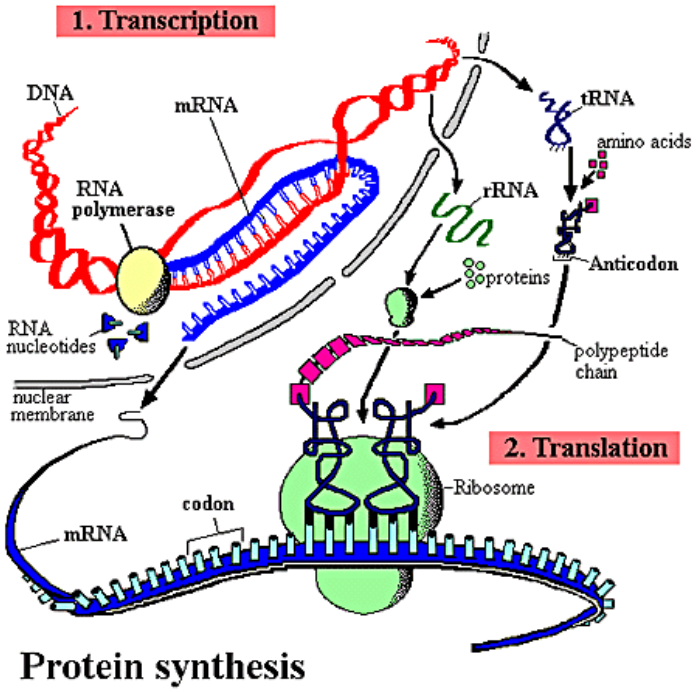

Figure 1.

In the case of some mechanisms, however, not only do the parts undergo changes but *which parts interact with which other parts changes* as the mechanism functions. The cyanobacterial circadian clock provides a relatively simple example. The core mechanism involves just three proteins, KaiA, KaiB, and KaiC and a source of ATP. The ATP provides phosphates that reversibly phosphorylate KaiC at two binding sites. Since one site is both phosphorylated and dephosphorylated first, the relative concentrations of KaiC in the different phosphorylation states uniquely specifies the time of day (Rust, Markson, Lane et al., 2007). (Note that in this case as well it is concentrations that do the work in the mechanism.) KaiC is itself capable of both autophosphorylation and

autodephosphorylation, and which operation it performs depends on KaiA and KaiB. KaiA can bind to KaiC in two different regions. When it acts alone, it binds to the A-loops coming out of the C2 domain of KaiC and fosters phosphorylation by changing KaiC's conformation (Figure 2, left). But when KaiB binds to the C1 domain, KaiA moves to bind to KaiB (Figure 2, right). By changing KaiC's conformation in a different manner, KaiA and KaiB promote dephosphorylation of KaiC (Tseng, Chang, Bravo et al., 2014). Even in this simple mechanism, KaiA interacts with different entities at different times, altering how the mechanism behaves. Which organization is implemented at a given time determines how still other parts (KaiC) behave and hence what time the clock registers. To exhibit this change in organization in diagram form, researchers often use separate diagrams to show the organization at different times.

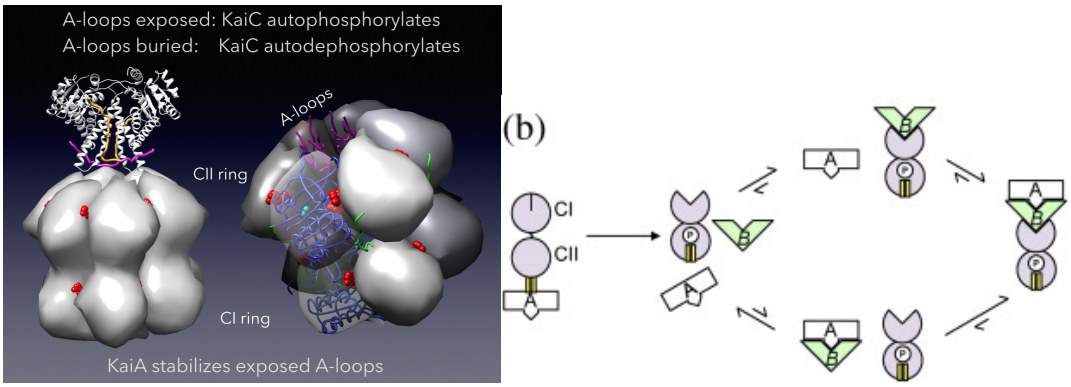

Figure 2. Left from Andy LiWang (http://faculty.ucmerced.edu/aliwang/galleries/image-gallery). Right: From Tseng, et al. (2014).

Examples such as this make clear that organization is a dynamic property of some mechanisms. The arrangement of parts, who interacts with whom when, varies as the mechanism functions, in part due to the operations the parts themselves are performing. So here we have another departure from Mechanism 1.0, inasmuch as it assumed a stable organizational pattern for mechanisms.

2.3. Mechanisms whose parts change over time

The examples we looked at so far involved changing concentrations and organizational patterns. But the "list" of parts playing a role in the mechanism was presumed stable. This is what we expect, based on the analogy to machines. But in some biological mechanisms, the parts change over time. Researchers have often failed to notice this since until recently it was not common to investigate a mechanism at different times. But as a consequence of automated data collection techniques, it has become possible to collect data about parts and their interactions at different times. For instance, in a study combining data about which proteins that can interact with each other to form complexes with time-series data on gene expression in yeast, de Lichtenberg, Jensen, Brunak et al. (2005) were able to provide evidence of how different parts are incorporated into a mechanism during different stages of the cell cycle. Although many genes are constitutively expressed, they

identified 600 genes (out of the approximately 6000 genes in yeast) that are only expressed during one stage of the cell cycle. Figure 3 shows one mechanism they investigated, the prereplication complex, which had previously been shown to involve Cdc28p and several Clb-type cyclins that function in regulating stages of the cell cycle. de Lichtenberg et al. demonstrated that individual cyclins are expressed and become available to bind with Cdc28p at different phases of the cycle. The color in Figure 3 shows the phase of the cycle in which the cyclins are synthesized: those shown in purple are expressed at the beginning of the M (mitosis) stage, those in orange through yellow during the G1 (gap 1) stage, those in green during the S (synthesis) stage, and those in blue during the G2 (gap 2) phase. At the end of the G2 phase the action of Cdh1p leads to the ubiquitination and degradation of the cyclins via Clb2p. The discovery that different parts are added to the mechanism at different stages of the cell cycle explains the different regulatory roles the mechanism plays at different stages.

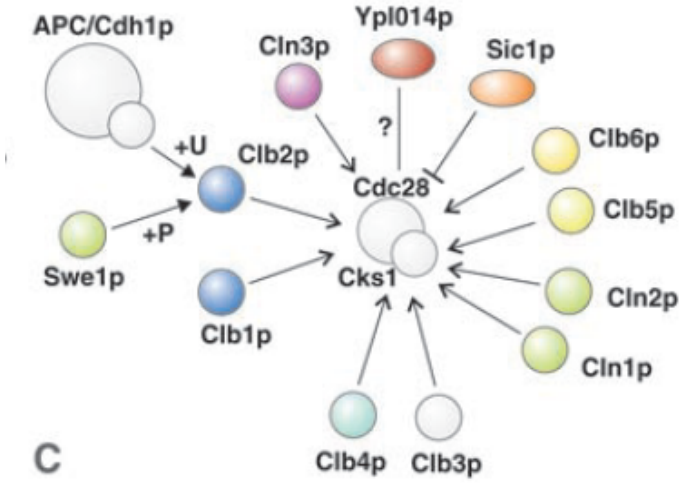

Figure 3. From de Lichtenberg et al. (2005).

The fact that the proteins that constitute the parts of some mechanisms change over time is not surprising. Biological mechanisms continually degrade and so are continually being built and repaired through the synthesis of new proteins. By not attending to when different parts are synthesized, Mechanism 1.0 tended to view mechanisms as enduring entities. Emphasizing their construction and degradation is thus a step beyond Mechanism 1.0.

2.4. Mechanisms with porous boundaries

When one buys a machine such as a toaster, it typically comes packaged in a box. The toaster cannot toast bread without electricity and bread being supplied, but the boundaries between the mechanism and the environment remain well delineated. Moreover, it operates in sharply distinguished time periods. Mechanism 1.0 made similar assumptions about biological mechanisms. An important discovery strategy was to try to localize the mechanism responsible for a phenomenon in time and space. These efforts often appeared

to be successful. Mammalian circadian researchers, for instance, introduced the term *clock* to designate the responsible mechanism and localized it initially in the suprachiasmatic nucleus of the SCN (Moore & Eichler, 1972) and subsequently in individual cells in the SCN (Welsh, Logothetis, Meister et al., 1995).[2] Treating circadian rhythms as produced within individual cells, investigators sought the component genes and proteins that constituted a feedback loops that created oscillations--a proteins accumulated, they fed back to inhibit their own expression (Reppert & Weaver, 2001) (Reppert and Weaver, 2001).

Researchers of course expected different mechanisms to have outputs that were used in other mechanisms--protein synthesis generated proteins that provide parts of other mechanisms. But over time researchers have found more and more interactions between mechanisms, so many that the notion of a mechanism as a well-delineated "thing", with specified boundaries in time and space, can come to look questionable.

We start with a specific example of an unexpected connection between mechanisms. As researchers investigated how circadian proteins fed back to alter the expression of their own genes, they discovered that one critical protein, CLOCK, affects gene expression as a histone acetyltransferase. In searching for a histone deacetylase needed to counterbalance CLOCK, researchers identified SIRT1, a molecule already known to be critical for a host of cellular activities including basic metabolism (Sahar & Sassone-Corsi, 2009; Bass & Takahashi, 2010) (left side, Figure 5). Parts identified as belonging to the mechanisms responsible for these activities also affect circadian rhythms.

As a result of sharing components, mechanisms affect the functioning of other mechanisms not just through their inputs and outputs, as characteristic of the examples advanced by Mechanism 1.0, but through many of their internal operations. The example of SIRT1 is just one of a host of discoveries researchers have made where parts of the clock mechanism interact with parts of other mechanisms involved in other cell functions. Using sRNA screens to identify genes that when modified affected clock performance, Zhang, Liu, Hirota et al. (2009) found many such genes involved in a wide array of other cell functions. On the right in Figure 5 core clock genes are shown in dark and light blue; those shown as connected in various ways to the core clock genes are the ones that altered the period of the clock when mutated and whose proteins are known to interact with core clock proteins. These are normally identified as parts of different cellular mechanisms.

[2] In the spirit of our discussion of concentrations above, registration of time of day appears to be a population level effect, but with the extra complication that there are mechanisms to promote synchrony in local populations and complex dynamics over the whole (Welsh, Takahashi, & Kay, 2010).

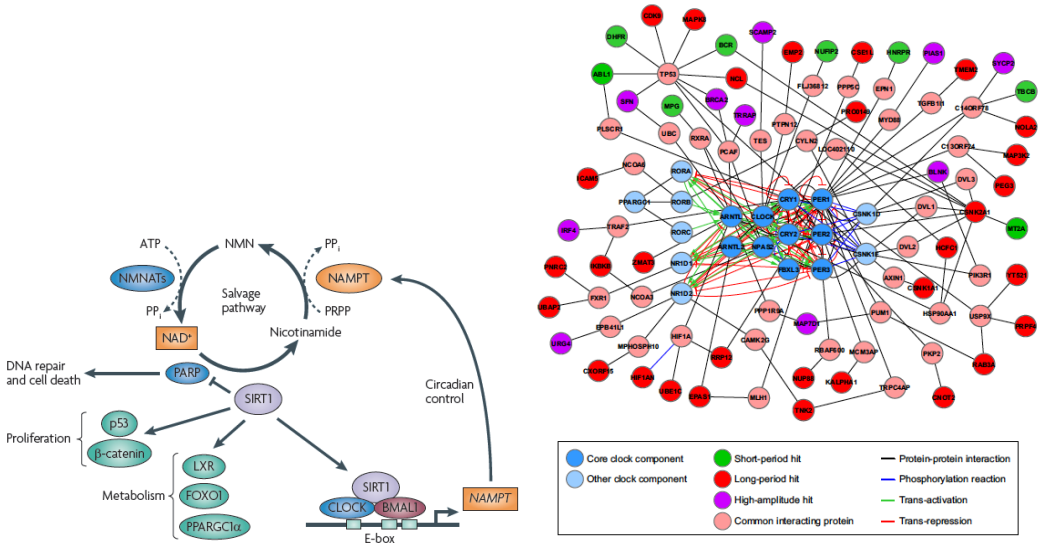

Figure 5. Left: Bass and Takahashi (2010). Right: Zhang et al. (2009)

These developments in circadian research exemplify a frequent trajectory of mechanistic research: in the wake of identifying a few parts of a mechanism, researchers continue to discover more and more entities that affect the functioning of the mechanism but are also recognized parts of other mechanisms. Recent efforts to represent the components of cells in networks (e.g., of protein interactions) often reveal that classically characterized mechanisms correspond to what are termed *modules*--clusters of parts that are more interconnected with each other than with other components, but which still have extensive connections to other modules. There typically are not clear boundaries around modules. Rather than finding sharp boundaries, researchers rather make pragmatic judgments as to where to draw boundaries (Bechtel, 2015).

The lack of sharp boundaries also affects the time window in which a mechanism carries out an operation. In standard portrayals of an action potential, the cell is at its resting potential until a stimulus arrives which depolarizes the cell. If depolarization exceeds a threshold, an action potential is generated. A recovery period follows in which the cell is first depolarized beyond the resting potential until it gradually returns to the resting potential where it resides until the next stimulus arrives. But in fact neurons fluctuate around the resting potential. Moreover, the effects of a single action potential can be demonstrated to affect these fluctuations, and hence the propensity to generate further action potentials up to minutes later, well after the action potential to reach its termination conditions (and may well have produced multiple additional action potentials). Marom (2010) argues that this shows that there is no characteristic timescale for action potentials. Nonetheless, researchers assume a timeframe for action potentials, ignoring these long transient effects.

These examples show that rather than well-delineated spatial and temporal boundaries between what is in and what is outside a mechanism, as are found with machines and is

suggested by the examples advanced by Mechanism 1.0, biological mechanisms often have porous boundaries. As a result of treating these boundaries as more fixed than they are, researchers are only able to account for phenomena approximately. When these shortcomings become important, researchers expand boundaries to include other parts and operations, but without abandoning the search for mechanistic explanation one cannot include everything. Selecting boundaries between multiple plausible candidates is an important challenge in Mechanism 2.0.

2.5. Mechanisms that exist only transiently

In Mechanism 1.0 mechanisms are viewed as present ready to operate when appropriate conditions arise. This corresponds to how we typically view machines, although when we adopt a long enough time horizon we recognize that machines are built, maintained for a period, and eventually decay, at which point they may be recycled. This occurs in biology far more frequently than suggested by Mechanism 1.0. Some biological mechanisms appear only to come into existence under specific conditions and are degraded when no longer needed.

Ideker and Krogan (2012) characterized *differential network biology* as a strategy in which network representations of gene or protein interactions identified under different conditions are contrasted to identify modules in yeast cells that only appear transiently. Employing annotations such as those provided by Gene Ontology, these modules can often be linked to mechanisms as identified in traditional molecular biology. The result is the identification of mechanisms that arise only in some conditions, presumably ones in which the phenomenon for which they are responsible is required by the cell.

Employing a version of this strategy, Bandyopadhyay, Mehta, Kuo et al. (2010) compared gene interactions in yeast growing under unperturbed conditions with those in which methyl methanesulfonate (MMS), a DNA-alkylating agent, was added to the medium. For each condition they created an epistatic microarray profile (E-MAP) that identified pairs out of a set of 418 selected genes that interact when mutated (that is, together they have effects on colony growth different from the product of their individual effects, as would be expected if they did not interact). Interactions are viewed as indicators that the proteins coded by the genes operate together in a mechanism. Bandyopadhyah et al. identified 1905 interactions in the untreated and 2297 in the MMS condition. Most of the interactions were only found under one of these conditions. They then created a differential E-MAP by subtracting the E-MAP for one condition from that for the other. This revealed many interactions not found when analyzing the conditions individually. In particular, the comparison revealed many interactions between DNA damage-response genes, which suggests that the proteins from these gene work together in mechanisms that arise in the MMS condition. When the gene interactions were mapped onto protein interactions, the researchers identified the differentially active connections as connecting between protein complexes. They interpreted the protein complexes as stable structures that are differentially recruited into mechanisms when specific tasks need to be performed.

In machines, components that are not needed on a given occasion just sit idle. Given the ability of organisms to synthesize proteins as needed, it is not surprising that, unlike machines, biological mechanisms construct mechanisms as need and then degrade them when not needed.

3. Mechanism 2.0

3.1. Recap

We have characterized Mechanism 1.0 in terms of the examples that were advanced in philosophical discussions of mechanistic explanation, and the features highlighted in discussions of them—such as the types of parts organization employed, and the kinds of discovery strategies used to identify such mechanisms. Our aim has not been to challenge the explicit definitions of mechanism that advocates of mechanistic explanation have advanced. Our justification is that it was the examples themselves and the discussions that emerged from considering them that made mechanistic explanation appear as a compelling alternative to other accounts of explanation, and that informed the community's understanding of the contrast between mechanistic and other forms of explanation.

Pursuing the examples advanced for Mechanism 1.0 has been extremely fruitful. The examples had the virtue of being widely intelligible by philosophers without extensive background in biology. And they painted a sharp and, we think, justifiable, contrast with alternative forms of explanation, including law-based explanations common in physics and some other sciences. However, as we have tried to show, these examples do not reflect the full scope of mechanistic accounts offered in science. The examples presented a view that is simplified along several key dimensions: portraying mechanisms as having discrete and enduring parts, organized in relatively fixed ways, and clearly distinguished from the environment in which they operated.

We have put forward other examples that differ in important respects. In many the operation of the mechanism depends on concentrations, rather than discrete parts. In some cases the parts change over time. Moreover, we presented mechanisms whose organization changes as the mechanism functions and which bled into their environment (including other mechanisms) rather than being sharply distinguished from it. Finally, we identified mechanisms that are transient, not enduring. All told, this puts the machine image underlying Mechanism 1.0 under severe strain.

3.2. Moving beyond the traditional machine metaphor

The examples put forward on behalf of Mechanism 1.0 conformed closely to the picture of machines designed by humans. Thinking about them as machines facilitated discovering and reasoning about them. Like machines, they are assumed to be localized in their environments and to have parts that are identified in terms of their structure. Researchers expect to be able to trace activity through the mechanism as it generates the phenomenon. Their internal states may change in the course of processing, but the overall organization remains constant. Several mechanists sought to differentiate biological mechanisms from

machines, but given the examples advanced to support Mechanism 1.0, the difference between biological mechanisms and machines was not all that clear.

The departures from Mechanism 1.0 on which we have focused put even more strain on the machine metaphor. We need to recognize, however, that *machine* is also an evolving notion. Historically, those opposing mechanism—vitalists and holists—emphasized the differences between biological systems and the machines then present (Bechtel, 2016). Inspired in part by Bernard and Cannon, cyberneticists (Wiener, 1948) expanded on historical conceptions of machines by emphasizing such things as the potential for control provided by negative feedback. Negative feedback, and its capacity for facilitating not only control but also oscillatory behavior, was a major inspiration for Mechanism 1.1. Sustained oscillators that can synchronize with each other and be entrained to external oscillations, provided examples of systems that don't wait for input to initiate activity but are endogenously active.

Today our conception of machines continues to evolve as designers explore options to make physical devices behave in ways not conceived of in the past. The ability to control electrical activity in computers through software, including software that can be modified by the machine itself, has certainly fostered this expansion in the concept of a machine. An additional factor has been the application of ideas about organization discovered in biology to designing new machines. Some of these are ideas we have characterized as motivation for Mechanism 2.0; incorporating them into machines may once again reduce the gap between machines and mechanisms.

Regardless of whether machines continue to evolve to more resemble biological mechanisms, our focus is on how biologists are revising their conception of how mechanisms are structured and, especially, of how mechanistic explanations work. Expectations for a localized, graphically depicted system, whose workings can be tracked via mental rehearsal and where discovery consists in large part of "looking under the hood," have been altered. Many biologists now recognize that mechanisms that differ from those advanced for Mechanism 1.0 are both common and important, and that modeling them requires advanced, typically mathematical, methods that go beyond flowcharts and structural figures.

3.1. Next steps

Where does this leave us? First, insofar as we have focused on the examples offered, not the definitions advanced, we are not contesting the definitions of mechanism. The letter of these definitions may well be compatible with several of the examples we have advanced. We don't see much benefit in the project of defining mechanism. Glennan (in press) advances the notion of *minimal mechanism* as providing a common basis for various more specific conceptions of mechanism, and this may suffice. Speaking loosely, we can treat any explanation that appeals to underlying parts-and-organization as mechanistic. This rough way of pointing to the kinds of systems and explanations at issue is all that we need.

If we are not contesting the definitions, then why are we making so much out of the ways our examples differ from those we see as characteristic of Mechanism 1.0? For one thing, we think it is the examples and the way they have been discussed that have shaped our understanding of mechanistic explanation. But more crucially, the interest in mechanistic explanation is not focused on developing adequate definitions but on issues such as how mechanistic accounts explain, how they are discovered and evaluated, and the ways in which they get applied in scientific reasoning. On this score, the examples fitting Mechanism 1.0 pointed in particular directions—the focus was on a well-delineated set of entities that were organized in a stable manner, were demarcated from others, and endured. With these examples in mind, it was natural to pursue particular kinds of strategies of research. For instance, decomposition and localization—i.e. breaking down a system into its parts and identifying their structure and place in the mechanism's layout—were emphasized by Bechtel and Richardson (1993). Strategies for constructing mechanistic hypotheses also followed this pattern, For instance, in the examples illustrating Darden's (2006) and Craver and Darden's (2013) strategy of modular subassembly (i.e. hypothesizing that a mechanism is composed of modules known from other mechanisms, in an altered configuration), the parts are presented as discrete and only interacting via their inputs and outputs. In their strategy of forward/backward chaining, one uses information about early (or, correspondingly, late) stages in the mechanism's operation to sketch hypotheses about later (or, correspondingly, earlier) stages. It is unclear that such a strategy can succeed when the mechanisms does not have well-defined boundaries, so that the kinds of constraints placed by earlier stages on the process are very hard to pin down. Thus, these and similar research strategies largely rely on discrete and localized parts, sequential organization and well-demarcated boundaries. They are unlikely to succeed with the types of mechanisms we have drawn attention to.

The mechanisms we have advanced on behalf of Mechanism 2.0 require different approaches. In cases in which concentrations matter, researchers often need to measure concentrations and collect time series data to understand how they change. If the mechanism only exists in certain contexts or if organization changes in different circumstances, then researchers need to contextualize the study of the mechanism and use tools that allow them to discern which components and processes are active when and where. Differential network biology, briefly discussed above, represents one such strategy. One should expect that the parts and organization of a mechanism may change as conditions in which it operates change. If mechanisms are not sharply differentiated but bleed out into their environment, then researchers may need to make different choices about the identity of the mechanism on different occasions.

As we are not advocating definitions, in advancing the notion of Mechanism 2.0 we are not seeking new definitions. Rather, we are advocating a broader research agenda that seeks different ways in which mechanisms can depart from the examples advanced for Mechanism 1.0 and still count as mechanisms. The more important task is likely to be distinguishing different kinds of mechanisms, potentially generating a taxonomy of

mechanisms. One dimension in such a taxonomy might be whether parts are localized or not and whether it is the structure of the parts or their concentrations that matter. Another might be whether the organization is enduring, changes under endogenous control, or changes in different environments. A taxonomy is only valuable if the different types of mechanism that are distinguished matter for philosophical and/or scientific objectives. We have made some suggestions as to how the different departures from Mechanism 1.0 do matter for understanding how a mechanism serves to explain a phenomenon and how the mechanism is discovered.

A taxonomy of mechanisms would address the diversity of mechanism types. A further set of questions concern mechanism tokens, namely: what are the identity conditions for mechanisms. When can we say that the same mechanism has changed over time and when do we have a new mechanism? Under Mechanism 1.0 this question hardly ever arose, but once we have mechanisms with changing parts, shifting organization and fuzzy or non-existent boundaries, it is natural to wonder whether and when one can speak of a mechanism as persisting through time.

We cannot resolve this issue here, of course, and it may very well be that the answer can only be given as part of a broader story about objects, change and identity through time. But let us indicate three potential directions on might proceed to individuate mechanisms. First, one can move to an "ephemeral mechanism" outlook, i.e. accept that when components and/or organizational features change, as they often do, we no longer have the same underlying mechanism. This entails that one and the same phenomenon may often be underpinned by different mechanisms over time. This is a somewhat unintuitive idea, as it severs, or at least weakens, the link between mechanisms and phenomena. But perhaps it is correct and unproblematic. A second option is to identify mechanisms via the phenomena they explain. On this approach, so long as we have the same phenomenon we have the same mechanism. Of course, this raises questions about the identity of phenomena. Not much has been written on this. The only well-developed account, Kaiser and Krickel (2016), construes phenomena as "object-involving occurrences." As the name suggests, this account presupposes a notion of (biological) objects, and it is not clear that such a notion can be retained in light of the cases we have looked at. Finally, one can view Mechanism 2.0 in the context of process ontology, the idea that the biological world, perhaps the world at large, consists of processes rather than objects (i.e. a temporally extended, constantly changing, "stream" rather than stably structured "thing"). Dupré (2012, 2014) has recently been arguing for such a view, and the cases we have discussed may provide more grist for his processual mill. Some will find process ontology to be a radical and implausible viewpoint, although there are ways of making it compatible with an explanatory appeal to mechanisms, especially if, as we tend to think, explanations are to be seen in epistemic rather than ontic terms. Here we remain uncommitted, as our principal aim is to highlight the issue as a topic for further exploration.

In concluding, we should stress again why we treat the examples we have described as mechanistic explanations despite the fact that some would treat them as non-mechanistic.

We think it is both more useful philosophically and in better accord with scientific practice, to expand our perspective on the types of mechanism that occur in biology. In expanding the scope of what qualifies as a mechanism, however, we are not emptying the notion *mechanism* of content. For one thing, the contrast between mechanistic explanation and DN or other formalist views of explanation is retained. Explanation is still a matter of describing the causal underpinnings of a phenomenon, rather than embedding it in a formal deduction schema. But beyond that, there are several sorts of explanations that can be seen to be non-mechanistic: etiological explanations, which chart a causal process leading up to an event are one example, as well as teleological explanations, which describe the function of an object or feature. Arguably, so are mathematical explanations (Pincock, 2007; Lange, 2012), which account for a phenomenon in terms of a formal-mathematical properties instantiated by it.

Clearly, we have only begun the task of transitioning from Mechanism 1.0 to 2.0. If our suggestion that there are multiple dimensions in which recognizably mechanistic science departs from Mechanism 1.0, Mechanism 2.0 might not have a univocal characterization but offer a taxonomy. As with the initial articulation of mechanism 1.0, we expect this project to be driven by close attention to the explanations actually offered in science. We also expect it to be as fruitful philosophically. And, if successful, it might ultimately itself be found wanting, making way for Mechanism 3.0.

References

Bandyopadhyay, S., Mehta, M., Kuo, D., Sung, M. K., Chuang, R., Jaehnig, E. J., Bodenmiller, B., Licon, K., Copeland, W., Shales, M., Fiedler, D., Dutkowski, J., Guenole, A., van Attikum, H., Shokat, K. M., Kolodner, R. D., Huh, W. K., Aebersold, R., Keogh, M. C., Krogan, N. J., & Ideker, T. (2010). Rewiring of genetic networks in response to DNA damage. *Science, 330*, 1385-1389.

Bass, J., & Takahashi, J. S. (2010). Circadian integration of metabolism and energetics. *Science, 330*, 1349-1354.

Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge: Cambridge University Press.

Bechtel, W. (2015). Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*.

Bechtel, W. (2016). Mechanists must be holists too! Perspectives from circadian biology. *Journal of the History of Biology*, 1-27.

Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A, 41*, 321-333.

Bechtel, W., & Richardson, R. C. (1993/2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Cambridge, MA: MIT Press. 1993 edition published by Princeton University Press.

Brigandt, I. (2013). Systems biology and the integration of mechanistic explanation and mathematical explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences, 44*, 477-492.

Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. Chicago: University of Chicago Press.

Darden, L. (2006). *Reasoning in biological discoveries: Essays on mechanisms, interfield relations, and anomaly resolution*. Cambridge: Cambridge University Press.

de Lichtenberg, U., Jensen, L. J., Brunak, S., & Bork, P. (2005). Dynamic complex formation during the yeast cell cycle. *Science, 307*, 724-727.

Dupré, J. (2012). *Processes of life: Essays in the philosophy of biology*. Oxford ; New York: Oxford University Press.

Dupré, J. (2014). A process ontology for biology. *The Philosophers' Magazine, 67*, 81-88.

Glennan, S. (in press). *The new mechanical philosophy*. Oxford: Oxford University Press.

Ideker, T., & Krogan, Nevan J. (2012). Differential network biology. *Molecular Systems Biology, 8*, 565.

Kaiser, M. I., & Krickel, B. (2016). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science*.

Lange, M. (2012). What makes a scientific explanation distinctively mathematical? *The British Journal for the Philosophy of Science*.

Levy, A. (2013). What was Hodgkin and Huxley's Achievement? *The British Journal for the Philosophy of Science*.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1-25.

Marom, S. (2010). Neural timescales or lack thereof. *Progress in Neurobiology, 90*, 16-28.

Moore, R. Y., & Eichler, V. B. (1972). Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. *Brain Research, 42*, 201-206.

Nathan, M. J. (2014). Causation by concentration. *British Journal for the Philosophy of Science, 65*, 191-212.

Nelson, P. C. (2015). *Physical models of living systems*. New York, NY: W.H. Freeman & Company.

Pincock, C. (2007). A role for mathematics in the physical sciences. *Nous, 41*, 253-275.

Reppert, S. M., & Weaver, D. R. (2001). Molecular analyses of mammalian circadian rhythms. *Annual Review of Physiology, 63*, 647-676.

Rust, M. J., Markson, J. S., Lane, W. S., Fisher, D. S., & O'Shea, E. K. (2007). Ordered phosphorylation governs oscillation of a three-protein circadian clock. *Science, 318*, 809-812.

Sahar, S., & Sassone-Corsi, P. (2009). Metabolism and cancer: the circadian clock connection. *Nature Reviews Cancer, 9*, 886-896.

Tseng, R., Chang, Y. G., Bravo, I., Latham, R., Chaudhary, A., Kuo, N. W., & LiWang, A. (2014). Cooperative KaiA-KaiB-KaiC interactions affect KaiB/SasA competition in the circadian clock of Cyanobacteria. *Journal of Molecular Biology, 426*, 389-402.

Welsh, D. K., Logothetis, D. E., Meister, M., & Reppert, S. M. (1995). Individual neurons dissociated from rat suprachiasmatic nucleus express independently phased circadian firing rhythms. *Neuron, 14*, 697-706.

Welsh, D. K., Takahashi, J. S., & Kay, S. A. (2010). Suprachiasmatic nucleus: Cell autonomy and network properties. *Annual Review of Physiology, 72*.

Wiener, N. (1948). *Cybernetics: Or, control and communication in the animal and the machine*. New York: Wiley.

Zhang, E. E., Liu, A. C., Hirota, T., Miraglia, L. J., Welch, G., Pongsawakul, P. Y., Liu, X., Atwood, A., Huss, J. W., Janes, J., Su, A. I., Hogenesch, J. B., & Kay, S. A. (2009). A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell, 139*, 199-210.