Rethinking Causality in Neural Mechanisms:

Constraints and Control

Jason Winning and William Bechtel
University of California, San Diego

1. Introduction

The nervous system in animals contains a multitude of control mechanisms that use information from various sources to regulate muscles and other tissues. The most basic level of control[1] is exhibited in reflex arcs, which in the simplest case involves a synapse between a sensory and motor neuron. In organisms with brains one finds local hierarchies of control, with higher-level controllers exercising their effects on more specific, lower-level ganglia and nuclei. Lower levels of control are often revealed when researchers remove higher-level control mechanisms (e.g., the cortex) and observe what control the organism still has over muscle movements. Comparable reliance on lower-level control systems is familiar to all of us who have, for example, navigated a highway while attending to something else and have no awareness of taking various actions.

As important as control mechanisms are in neuroscience, they are not afforded a central place in the main philosophical approach to explanation in neuroscience—mechanistic explanation.[2] Mechanistic explanations are characterized as starting with a delineated phenomenon (e.g., the generation of action potentials, the perception of objects, or the encoding of episodic memories), and then proceed by identifying the responsible mechanism, decomposing it into its parts and operations, and showing how, when the parts and operations are appropriately organized, they together produce the phenomenon (Bechtel and Richardson, 1993/2010; Bechtel and Abrahamsen, 2005; Glennan, 1996 and in press; Machamer, Darden, and Craver, 2000; Craver, 2007; Craver and Darden, 2013). As Glennan makes explicit, the phenomena are causal processes as are the operations appealed to in explaining them. The explanation involves decomposing one causal process into an organized set of causal processes that are thought to be responsible for it.

---

[1] The use of *levels* in the context of control is distinct from the notion of levels of organization or levels in a mechanism (as discussed by Craver, 2007). Although control can be exercised by a mechanism on its component parts, control can also be exercised by completely separate mechanisms or even by parts of a given mechanism.

[2] Philosophers presenting mechanistic explanation have discussed examples of control mechanisms, such as negative feedback (Bechtel, 2011; Bechtel and Abrahamsen, 2011) and circadian mechanisms (Bechtel, 2010; 2013), but they have said little about how these mechanisms effect control on other mechanisms. The analysis presented in this paper is intended to fill that lacuna.

What does "responsible for" mean here? Mechanist philosophers have construed this relationship in terms of either composition or causation (i.e., causal *production*). But there is a third type of relationship that often exists between processes in mechanisms—*control*—which has important implications for understanding the nature of mechanisms and mechanistic explanation that have not yet been fully appreciated.[3] Control relationships within mechanisms give rise to dimensions of organization that are missed by existing accounts of mechanistic explanation, which only focus on how mechanisms are organized in terms of composition and production. By focusing on a given causal process as the phenomenon, marshaling the parts and operations that together produce the phenomenon and, as Machamer, Darden, and Craver do, construing mechanisms as working from start to termination conditions, mechanistic accounts do not treat production as distinct from control.[4] In this paper, we advance a substantial revision to the traditional mechanistic perspective in which processes are controlled by other processes, and mechanisms are controlled by other mechanisms, often hierarchically.

We develop our framework by considering three other shortcomings of mechanistic accounts of explanation. As we noted, a mechanistic explanation accounts for one causal process in terms of the causal processes of the mechanism's components. But as Machamer, Darden, and Craver argue, this strategy cannot be iterated indefinitely; at some point this process must *bottom out*.[5] Lacking a metaphysical foundation to end the regress, one that at some point explains causation without resorting to the next mechanistic level down, leaves a cause having a given effect as simply a brute, unexplained fact. We term this the *mysteriousness* problem. As Kuhlmann and Glennan (2014) argue, further decomposition of the mechanism into other entities and activities at best only pushes the mystery deeper. It doesn't resolve it. Rather, it identifies other brute, unexplained causal relations. This inability to explain the specific effects a cause has is not helped by adopting Woodward's (2003) view of causation, as some mechanists (e.g., Craver, 2007) have done. Woodward himself has conceded that his view is intended merely as an account of "how we think about, learn about, and reason with various

[3] We are not suggesting that control is non-causal, but that it is something more than merely one process causally producing another.

[4] In defending top-down causation, Craver and Bechtel (2007) argue that conditions that affect whole mechanisms also affect their parts, thereby providing a sense in which activities of whole mechanisms control those of their constituents. More recently, Bechtel (2017) has further characterized top-down causation in terms of activity in a larger system imposing constraints on individual units in the network. While this proposal resembles in some respects the one we advance here, it is limited to the context of top-down causation, while the account we offer applies more generally to cases in which one mechanism exercises control over another.

[5] For that matter, the causal efficacy of the whole mechanism remains mysterious as well. Machamer, Darden, and Craver (2000) appeal to the productive continuity from one activity to the next and Bechtel and Abrahamsen (2005) appeal to the ability of researchers to simulate mentally the component operations to show how they generate the overall phenomenon. But actual accounts of mechanism are typically incomplete, and the gaps in the mechanism are sometimes only revealed much later after the explanation has been widely accepted.

causal notions and about their role in causal explanation" (2008, p. 194); he explicitly denies that his account is intended to solve the sort of bottoming-out problem that mechanists have been concerned with.[6]

The second and third shortcomings on which we focus are seldom discussed in the mechanist literature. The second is that biological organisms are "dissipative structures," meaning that they actually use the second law of thermodynamics to their advantage to maintain their organization. Organisms are able to extract the energy they need to do the work of self-maintenance, survival, and reproduction from their environment. Unlike most things, organisms actually thrive instead of degrade over time in the presence of dissipative energy flows in the environment and within themselves. An account of the mechanistic nature of biological systems should be able to account for how organisms manage to extract the energy they need from their surroundings (the *dissipative structure* problem). Such an account should also explain how such energy flows are channeled so as to contribute to the organism's own organization and functioning. This leads to the third problem: organisms must control (also build and maintain) themselves and procure their own energy—the *biological autonomy* problem.

Although these problems may seem to be independent, they are related by having a common solution. The key to solving all three is found in a surprising field: classical mechanics. To account for the behavior of macroscopic objects, theorists in classical mechanics introduced the notion of *constraints*.[7] Constraints supplement fundamental force laws, limiting the degrees of freedom available to elementary particles, and determining that macroscopic objects behave in specifiable ways. A number of theoretical biologists, many inspired by the pioneering work of biophysicist Howard Pattee, showed how an understanding of constraints can provide critical insights into biological systems. Philosophers of biology, including the proponents of mechanistic explanation, have largely ignored their work. We will argue that by extracting the ignored insights of theoretical biologists, we can provide a more adequate account of causality in biological mechanisms, one that answers the three problems identified above and makes clear how mechanisms are different in crucial ways from human-made machines. This conception of causality in biological mechanisms will then provide a framework for conceiving of them as regulated by control mechanisms, including neural mechanisms.

[6] See Glennan (2009) for an argument to the effect that the virtues of Woodward's account of the epistemology of causal reasoning "[do] not legitimate the manipulability theory as a metaphysical account of causation" (2009, p. 318).

[7] See Kuhlmann and Glennan (2014) for further discussion about how the classical mechanical causation of macro-level mechanisms can be understood as compatible with quantum mechanics on the Copenhagen interpretation. Rather than seeing theirs as a competing account, we believe that their paper fits well with this account because it is essentially a discussion about the relationship between classical and non-classical types of constraints.

We begin in the next section by introducing the notion of a constraint and using it to lay the foundation for a non-Humean, causal powers-based account of mechanistic causal production. This provides an answer to the mysteriousness problem. Then in sections 3 and 4, we address the dissipative structure problem and biological autonomy problem, showing first how constraints are critical to a machine or mechanism performing work, and then showing that certain kinds of constraints provide the means of maintaining organisms far from equilibrium. Given variable environments, the flow of energy must be controlled and the behavior of the whole system[8] controlled so as to procure energy. For a biological system to be autonomous, all control must be exercised by mechanisms within it. In section 5 we develop a number of examples of control systems found in human-built machines and in biology. Finally, in section 6 we apply this framework to the control exercised by the nervous system and the brain, which are often organized into local hierarchies of control mechanisms.

2. Powers, Constraints, and Mechanistic Causation

2.1. The Mysteriousness Problem

The most common way of providing an account of the metaphysical grounding of causation in philosophy has been to appeal to laws. But as has been argued by many authors (e.g., Smart, 1963, pp. 50–61; Bechtel & Abrahamsen, 2005, p. 422), law-based accounts do a poor job of characterizing causation in biology. Machamer, Darden, and Craver have argued that a suitable metaphysical account of mechanistic causation should make the principle of change *intrinsic* to the mechanism itself, rather than locate it extrinsically in some external metaphysical entities such as laws. Their move is to appeal to *activities* as a fundamental ontological category and to argue that this provides an account of what it is *in virtue of which* changes in mechanisms occur:

> [I]t is artificial and impoverished to describe mechanisms solely in terms of entities, properties, interactions, inputs-outputs, and state changes over time. Mechanisms do things. They are active and so ought to be described in terms of the activities of their entities, not merely in terms of changes in their properties. (2000, 5)

This move, however, is insufficient to address the mysteriousness problem. Machamer, Darden, and Craver are right to emphasize that mechanisms are intrinsically active in the sense that mechanistic explanation appeals to the causal efficacy inherent to mechanisms themselves, not

[8] It has been common to focus on organisms, especially single-celled organisms, as the locus of biological autonomy (Moreno & Mossio, 2015). However, many organisms live in symbiotic relations in which crucial activities are shared between numerous organisms, often from multiple species (O'Malley, 2014). Control relations such as we discuss later can involve entities in the environment with which an organism is tightly coupled. Accordingly, when considering autonomy, we speak of *biological systems*, not *organisms*.

to an external agent or extrinsic principle of change. Activity considered apart from the principle of change that brings it about is nothing more than a change in properties, which could be brought about by laws, powers, capacities, a deity, an Aristotelian "unmoved mover," etc. Simply citing the fact that there is activity does not account for the sense in which the activity is intrinsic versus extrinsic. Laws, deities, reified counterfactuals (whatever that would amount to), and unmoved movers are examples of extrinsic principles of change. Powers, natures, and capacities are examples of intrinsic principles of change. The activity of a mechanistic component is *what* the component actually does; it is not the *why*. For this reason, we contend that what is needed is an account, not of the *activities* of mechanisms, but instead of whatever it is about mechanisms a) that is intrinsic to them, and b) in virtue of which activity is brought about.

Philosophers of science have been slowly warming up to the idea of explaining mechanistic causation in terms of causal powers.[9] Such a move goes against the instincts of many philosophers of biology, who are especially wary of non-Humean metaphysical entities since the failure of vitalism. Machamer (2004), for example, contends that if mechanists appealed to powers or abilities, this would leave mechanistic causation metaphysically *mysterious*:

> activities are better off ontologically than some people's ontic commitments to capacities, dispositions, tendencies, propensities, powers, or endeavours. All these concepts are derivative from activities. ... [T]he active exercise of a capacity has to be ontologically prior to any mysterious property called "the ability to exercise that capacity." (2004, p. 30)

We contend that Machamer has it exactly backwards. An activity cannot be the exercise *of a capacity* if the capacity consists of nothing over and above activity. A capacity or power is *explanatory* precisely because it is what gives rise to the activity. In order to do that, it cannot be ontologically reducible to the activity. Further, Nancy Cartwright warns that it is important to avoid "conflation of the manifestation, or exercise, of [a] capacity with the occurrence of the canonical behavior we associate with the capacity" (2008, p. 195).

Part of why philosophers like Machamer and Craver have preferred activities over something like powers (and perhaps why Machamer thinks powers are "mysterious") is because such modal, non-Humean notions do not seem to be part of the conceptual lexicon of science: scientists don't tend to talk in a way that makes it obvious that they are committed to the reality of entities like powers. Cartwright has argued that a commitment to capacities or powers is "implicit … in the conventional methods for causal inference" (1989, p. 142) used by

[9] An approach to mechanistic causation along similar Aristotelian lines was defended recently by Cartwright and Pemberton (2013).

scientists, but for those who are unconvinced, we offer a different type of argument. We argue that the conceptual lexicon of science *does* have the resources to solve the mysteriousness problem and explain how causal production is grounded in mechanisms. Specifically, philosophers have mostly overlooked a modal conception that has been key to physics since the eighteenth century: *constraint*.

2.2. Constraints

The notion of a constraint has its roots in classical mechanics where it was introduced to address the challenge of explaining the behavior of macro-scale objects (Sklar, 2013; Hooker, 2013). Newtonian force laws can be applied to any physical particle to determine how it will behave in response to forces imposed upon it. Each particle has six degrees of freedom (along the three spatial dimensions and around each spatial axis). To use force laws to determine the behavior of a particle in response to a force imposed on it, one must calculate the change in values of the variables for each degree of freedom. When these particles are assembled into macro-scale objects, one can still calculate the six variables for each particle, but that soon becomes an overwhelming task. It also proves to be unnecessary. What makes a set of particles into a macro-scale object is a set of constraints that fixes relations between possible values on different degrees of freedom and so restricts the possible trajectories along which particles can move. For example, if two molecules form a rigid bond, they will be displaced together and it suffices to consider only one variable for each dimension for the combined object. Constraints operate not just within but also between macro-scale objects: when a marble, held together by bonds that limit the degrees of freedom of the atoms that constitute it, rolls on an inclined plane, itself a macro-scale object constituted by bonds that constrain the freedom of the atoms constituting it, the plane constrains the marble, and hence further constrains the movement of its particles.

By introducing a vocabulary that designates macro-scale objects, researchers can formulate rules or laws that characterize the behavior of macro-scale objects. But these laws are not the same as the force laws describing the movement of the particles and they cannot be derived from the force laws alone. This is because the force laws alone do not determine the constraints. Rather, constraints must be identified empirically and constitute additions to the representation of the force laws (e.g., to the Lagrangian/Hamiltonian formalism). In the parlance of theory reduction framed in terms of deductive-nomological accounts of explanation (Nagel, 1961), constraints are boundary conditions that allow deriving the laws of the reduced science from those of the reducing science. While the reduction framework emphasizes the linkage of reduced to reducing sciences, the reliance on boundary conditions, which must be determined empirically, renders the science concerned with macro-scale objects semi-autonomous. As a result of incorporating constraints, scientists can develop generalized

accounts for the interactions of macro-scale objects that ignore the degrees of freedom that are foreclosed when the constituents are incorporated into the macro-scale objects. The framework of constraints can be applied iteratively—a macro-scale object can be further constrained by incorporating it into a yet larger-scale object. For example, the ways in which a macromolecule might move are further constrained when it is embedded within a membrane of a cell, and that membrane is further constrained when the cell is incorporated within a multicellular organism.

2.3. How Constraints Ground Mechanistic Powers

A mechanism is, among other things, a type of dynamical system.[10] To appreciate how constraints ground causal powers, it is useful to employ the framework of state-space representations from dynamical systems theory in which there is a dimension for each variable characterizing a system. Thus, any possible state of the system corresponds to a point and a change in the system is represented as a trajectory through this high-dimensional space. What constraints do is restrict trajectories from reaching some parts of the state space. For example, if two particles are rigidly bound, then the system cannot reach points in the state space representing the particles having distant locations. But constraints also bias a constrained object towards reaching points and trajectories in the state space that would otherwise have been practically impossible or vanishingly unlikely.

A completely unconstrained system will have no behaviors; it would simply be disorganized motion of particles. Mechanistic causation consists in the causal production of changes in virtue of the organization of the system; in Kuhlmann and Glennan's words, "While mechanists emphasize the importance of spatial and temporal organization, it is ultimately the causal organization upon which the productive capacities of the mechanism depend" (2014, p. 341). We argue that the *causal organization* of a system consists exactly in its spatiotemporal organization combined with the operative constraints.

Take atoms and molecules. These have the causal characteristics that they have in virtue of their nuclear and electromagnetic bonds; they are not simply an aggregation of protons, neutrons, and electrons. A bond is a constraint. Proteins, the molecules within biological mechanisms that perform most of the work, have the ability to do the work they do in virtue of how they are constrained, i.e., their primary, secondary, tertiary, and quaternary structures. It is these constraints, including phenomena such as covalent, ionic, and hydrogen bonds, hydrophobic interactions, and van der Waals forces, that reshape the trajectory landscape of

[10] For a mechanism to produce a phenomenon it must undergo changes induced by the activities of its parts and so fits the broad conception of a 'dynamical system' as "a structure of mutually and simultaneously influencing *change*" unfolding in real time (van Gelder & Port, 1995, p. 3).

molecules and aggregations of molecules, and combine into higher levels of organization to yield complex systems with novel, emergent behaviors.

This might seem counterintuitive and paradoxical: how can something like a constraint, that is a *limitation* on behaviors, be a factor that *enables* behaviors? Consider an example from Cliff Hooker:

> a skeleton is a disabling constraint, for example limiting the movements of limbs (cf. an octopus), but by providing a jointed frame of rigid components for muscular attachments it also acts to enable a huge range of articulated motions and leverages, transforming an organism's accessible niche, initiating armour and predator/prey races, and so on. (2013, p. 761)

Each protein in an organism is like a skeleton in this sense; a structure that will resist certain forces while translating the directions of other forces, re-routing forces, displacing forces, etc. all by virtue of how it is constrained. The tendency or capacity to resist, re-route, displace, etc. various forces is just what it is to be a causal power. Thus, on our view, when constraints enable objects to have novel, emergent behaviors, this is tantamount to the emergence of causal powers. Enzymes provide an exemplar of how constraints generally account for the causal activities of mechanisms. The ways that mechanisms and their parts are constrained explains why both mechanisms and their components are intrinsically active; by means of possessing such emergent powers, mechanisms and components causally produce the effects they do.

The constraints realized in skeletons and proteins determine the possible behaviors of those objects. That is, they determine modal features—what these objects can and cannot do. It might be objected at this point that this modal nature represents a lingering thread of mysteriousness in our account. Philosophers of science Don Ross and James Ladyman, who are about as naturally inclined as philosophers can be,[11] endorse the physical reality of primitively modal "locally dynamic real patterns," and they also use the word 'constraint' to refer to these.[12] For Ross and Ladyman, there are no more basic principles of change and modality beyond these local constraints; they, and not possible worlds, laws, or counterfactuals, are the ontological bedrock of dynamical organization. The causal power theorist can endorse this picture and hold that it is in virtue of constraints, so interpreted, that systems have causal

[11] See chapter 1 of their book, where they rail against "esoteric debates about substance, universals, identity, time, properties, and so on, which make little or no reference to science, and worse, which seem to presuppose that science must be irrelevant to their resolution" and the associated tendency to prioritize "armchair intuitions about the nature of the universe over scientific discoveries" and to attach "epistemic significance to metaphysical intuitions" (Ross, Ladyman, & Spurrett, 2007, p. 10).

[12] Other authors that seem to invoke something similar to this physical and modal interpretation of constraints are Esfeld (2009) and Kistler (2009).

powers, including their modal features.[13] The notion of a constraint can explain the intrinsic causal activeness of macro-scale objects, including biological mechanisms, in a way that avoids mysterious or "spooky" metaphysical baggage of the sort Ross and Ladyman would reject.

3. Work, Energy, and the Dissipative Structure Problem

Any macro-scale object exhibits constraints: crystals, solar systems, human-built machines, etc. What sets the constraints employed in machines and biological mechanisms apart is that they enable those systems to perform work—to act on and alter other systems (or, in more complex cases, parts of themselves) in a systematic way. Introducing work requires that we introduce the corollary notion of Gibbs free energy, the thermodynamic potential that determines that maximum work that can be performed. The concepts of work and energy have not figured in philosophical accounts of mechanism, but in this section we will show that they have important roles to play in characterizing both machines and biological mechanisms and in showing how biological mechanisms differ from human-built machines.

To begin with the commonality, in both machines and biological mechanisms, constraints serve to direct the flow of free energy in a coherent manner such that work can be performed. The importance of this can be appreciated if we first consider two constrained systems that are not appropriate to perform work. Crystals possess rigid constraints; their constraint structures do not allow their constituent molecules any freedom to move relative to one another. The crystal acts as a rigid body and might transmit a force imposed on it to something else, but it does not itself perform any work. At the other extreme, a rigid container constrains the movement of particles within it. The particles can move independently of each other, but are limited to moving within the container. The particles in the container cannot do work. Neither set of constraints is capable of directing the flow of free energy in the performance of work.

To perform work, as Atkins shows in discussing the Carnot engine, a system must be able to *selectively filter* the flow of energy so as to shape its response properties:

> Here is an essential asymmetry of the engine: it possesses a directional response to the impacts it receives. The face of the piston is, in effect, a screen: it picks out and responds to the motion of particles that happen to be traveling perpendicular to it; and it rejects (by not responding to) components of motion that happen to be parallel to it. Engines, in effect, select certain motions of the particles within them. The directionality of the movement of

[13] Juarrero (1999, pp. 131–132) and Moreno and Mossio (2015, p. 51) have also proposed that constraints can ground causal powers. For a more fully worked out version of the present account of causal powers, see first author (forthcoming).

> an actual piston in an engine is a consequence of this asymmetry. Our exploitation of heat to achieve work is based on the discovery that the randomness of thermal motion can be screened and sorted by asymmetry of response. (1984, p. 83)

The notion of constraints serving as filters clarifies respects in which machines are intrinsically passive and respects in which they are intrinsically active. They are intrinsically passive in that they must make use of free energy from the environment to perform work. They are intrinsically active in that they *filter and shape* the flows of free energy from the external and internal environment. *All* mechanistic operations are, among other things, *energy flow* operations filtered and shaped by constraints.

To understand how biological systems are organized to perform work, Pattee (1971) discusses two important distinctions about how constraints in dynamical systems are described mathematically in analytical mechanics. In some simple machines, such as the lever or the screw, the way the system is constrained does not change over time. The constraints function as time-invariant limitations on how energy is transferred and the affected objects move. These can be represented in terms of the coordinate variables in the equations describing the system, but time does not enter explicitly as a variable. Other simple machines, such as a system of pulleys, are usefully described as being constrained in different ways at different times. Their constraints might instead be mathematically represented by equations that include time as an explicit variable. The central point is that such constraints are time-dependent.[14]

A second important distinction is whether the constraint equations are integrable or not. Constraint equations often can be stated in terms of the variables in the equations describing the system or only in terms of their derivatives. In the former case, the motion of the system adheres to a fixed geometry, and is state-determined; as a result, the constraint reduces the number of variables (degrees of freedom) needed to describe how a system changes over time. In the latter case, on the other hand, the constraints limit the dynamics of the system in a way that is not geometrically fixed, so that it can still respond to perturbations in a flexible way. The former case corresponds to a dynamical system that changes over time but is rigid in that it has no "give" at all when perturbed by forces from outside the system.[15]

Many constraints in natural systems are time-dependent and can only be represented in terms of derivatives. They change in complex, non-linear ways—either as a result of work performed

[14] Time-invariant and time-dependent constraints are represented in analytical mechanics by scleronomic and rheonomic constraint equations, respectively.

[15] In analytical mechanics, integrable constraint equations, yielding a state-determined dynamics (e.g., particles of a rigid object, or a series of tightly intermeshing gears), are called *holonomic*, whereas non-integrable constraints, yielding a flexibly constrained system (e.g., particles free to move but confined within a box, or loosely intermeshing gears), are called *non-holonomic*.

within the system or by other systems external to it. For example, water flowing in a river is constrained by the banks but also alters the banks through erosion. Humans who seek to redirect the flow of water can also alter the banks. These constraints cannot be represented just in terms of a function of the coordinates of the particles but require specifying their velocities. Consequently, they are non-integrable and their dynamical profile changes as a function of time. Flexible time-dependent constraints are common in the biological world and we will focus on them in subsequent sections. But they are also common in human-built machines. Switches are an example. A train switch, which historically needed to be operated by the application of human work, directs a train along one track or another.

Existing mechanist approaches see biological mechanisms as analogous to human-built machines. An important disanalogy, however, arises when one considers how energy figures in machines versus biological mechanisms. Human-built machines can usually take their energy supply for granted. The human user will insert a battery, or charge the battery when it gets low, or put gas in the gas tank. Or the human user will wind up the machine, or situate it in a place where the machine has a steady stream of power coming into it (plugged into a wall socket, or underneath a waterfall, or in outer space with solar panels aimed at the sun) and equip it to take advantage of that power source. If the human fails to do this, no work will be performed. The situation is very different for biological systems; they do not have batteries, a 110 volt AC plug, or a wind-up handle. Nonetheless, the mechanist framework has simply assumed that the needed energy is available to biological mechanisms. The account of protein synthesis, analyzed by Darden and Craver (2002), for example, assumes a source of energy in ATP and does not address where it comes from or how it is employed in the work of synthesizing proteins. This is an important omission. In addition to accounting for parts, operations, and organization, biological explanations must also account for how energy gets to the parts, how it is transmitted between them and/or stored by them, and how it is utilized and dissipated in mechanistic operations (i.e., *work*). In biological systems, unlike artificial machines, these aspects are generally inseparable and are importantly shaped by one another; an account of how biological phenomena are produced must generally also be an account of the flow of energy, and vice versa.

Energy is needed not just for mechanisms to perform work, but also to maintain the mechanisms themselves. Biological mechanisms are dissipative structures—they occur and are maintained in contexts in which free energy is being dissipated (entropy is increasing). Prigogine emphasized that in order to be a *self-stable* dissipative structure (i.e., one that does not rely on human operators to continually supply energy to it), a system must be appropriately constrained (Nicolis & Prigogine, 1977). By imposing specific local constraints on the flow of free energy, dissipative structures can slow the process of moving towards equilibrium to a

crawl or even reverse the process in some locations at the cost of speeding it up in others (see Atkins, 1984, p. 108).

4. Control, Self-Regulation, and the Biological Autonomy Problem

In the previous section we argued that free energy and work are key concepts for an account of machines and biological mechanisms. Especially in the case of biological mechanisms, which are dissipative structures, constraints must play a critical role in procuring and directing the flow of free energy. Most human-built machines are designed to operate in situations in which energy is reliably supplied. This is typically not the case for biological systems. They have to be responsive to changing conditions: for example, routing free energy in different ways on different occasions and seeking a different form of free energy when the current one is no longer available. This requires exercising control over time-dependent, flexible constraints so that their component mechanisms perform the work required to maintain the whole set of mechanisms that constitutes the organism in a far-from-equilibrium condition. When this fails, the organism dies. With considerations such as these in mind, Ruiz-Mirazo and Moreno flesh out their conception of *biological autonomy*, which they take to be a key criterion for life, as

> the capacity of a system to manage the flow of matter and energy through it so that it can, at the same time, regulate, modify, and control: (i) internal self-constructive processes and (ii) processes of exchange with the environment. Thus, the system must be able to generate and regenerate all the constraints—including part of its boundary conditions—that define it as such, together with its own particular way of interacting with the environment (2004, p. 240; see also Ruiz-Mirazo, Peretó, & Moreno, 2004, p. 330; Varela, 1979).

Some of the basic requirements for control can be observed in human-built machines in which a human operator exercises control. An operator controls a machine's time-dependent, flexible constraints, thereby changing the work that is done. For example, the operator of a traditional car manipulates constraints to cause it to move forward or backwards, adjust its speed, or stop its motion. Altering the speed of a car involves modulating the flow of gasoline into the carburetor by altering a valve. This is done through a linkage mechanism that connects the gas pedal to the valve. By pressing or relaxing pressure on the pedal, the driver alters the constraints that determine the flow of gasoline into the engine.

In the car scenario, the driver is external to and exercising control on the car. In living systems, individual mechanisms may be controlled by other mechanisms, just as in the car case. But this control must be exercised from within the biological system. Biological systems are autonomous systems in the sense employed by Ruiz-Mirazo and Moreno above; they are not dependent on an external agent to control them.

We cannot here fully address the biological autonomy problem.[16] But we offer four observations about how control is typically realized in biological systems that have implications for how to conceptualize the connection between control and biological mechanisms. First, control is exercised on already functioning mechanisms that are capable of directing free energy into the performance of work. The controller does not need to fully direct the behavior of the controlled system but only to modify its operations. Typically, in fact, control involves down-regulating a capacity that would otherwise be executed. Second, control is often highly distributed among various mechanisms that each performs a different control function on the same basic mechanism. Third, control mechanisms themselves are often controlled by still other mechanisms within the same biological system. This typically involves another mechanism operating on a time-dependent, flexible constraint in the first control mechanism. Sometimes, but not always, these control mechanisms are organized hierarchically. As with basic control mechanisms, when the higher-level control mechanism is incapacitated, the lower-level control mechanism can continue to operate by virtue of its own constraint structure. But finally, there is no infinite regress up a control hierarchy. Rather than a strict hierarchy, controlled and controller mechanisms typically form interconnected networks in which control over mechanisms at the top of a local hierarchy may be exercised by yet other mechanisms in the biological system (realizing what Pattee, 1991, characterized as a *heterarchical* control model[17]).

To illustrate these aspects of how control is realized in biological systems, we first offer examples of control mechanisms that do not involve neurons or brains. We then turn in section 6 to neural control.

5. Feedback Control in Machines and Biological Mechanisms

At times, humans have built machines that implement the type of internal control biological mechanisms must employ to achieve autonomy. The governor James Watt designed for the steam engine is a classic example. The steam engine involves a set of constraints that directs steam energy to turn a flywheel (d in Figure 1) to which a number of appliances such as sewing machines are attached. The challenge Watt confronted was to prevent the flywheel from increasing or decreasing its speed as the number of appliances operating changed. This required altering the flow of steam through a valve (Z), a flexible time-dependent constraint in the engine itself. The governor Watt created involved a spindle (D) attached to the flywheel to which arms (E) were attached. Due to centrifugal force, these arms moved further out when

[16] In addition to explaining self control within biological systems, a full account of autonomy would require explaining how biological systems build (Varela, 1979) and repair (Rosen, 1991) themselves.

[17] On the notion of *heterarchy* and how it contrasts with *hierarchy*, see also McCulloch (1945), Turvey (1977), and Yates (1979).

the speed of the flywheel increased. He then attached a linkage mechanism (c, f, h, F, G, H) that closes the valve proportionately to how far out the arms extended. The governor implemented the design principle of negative feedback: as the flywheel rotated faster, the arms were thrown out further, which (via the linkage mechanism) would partly close the valve and thus reduce the flow of steam. Likewise, as the engine moved more slowly, the arms would drop, causing the valve to open further and speed up the engine. Depending on how the weights of the arms and the linkage mechanism were calibrated, there would be a target speed that the flywheel would approach (or oscillate around).

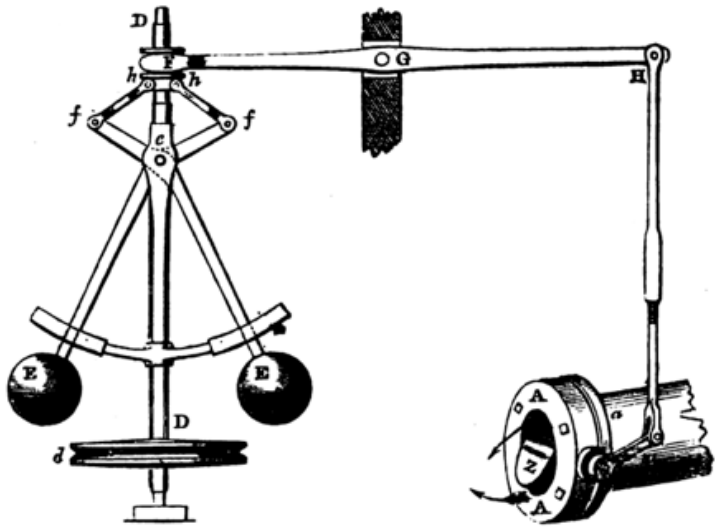


Figure 1. The centrifugal governor for the steam engine designed by Watt. See text for details.

The Watt governor is an example of control system employing flexible time-dependent constraints that one can understand intuitively: the angle of the angle arms appears to represent the speed of the flywheel. One can also analyze the engine, as Maxwell (1868) did, in terms of differential equations. When one solves these equations to show, for example, how the angle arms carry information about the speed of the flywheel, one discovers that this information is in fact represented by the angle, the velocity of the angle, and the acceleration of the angle (Nielsen, 2010). The fact that a representation of the dynamics must include as terms not only the positions of parts but also velocities and acceleration is characteristic of flexible time-dependent constraints (which Pattee, 1970 appropriately also refers to as ‘machine-like constraints’).

Negative feedback illustrates the idea that control does not directly require the intervention of an external human agent.[18] Although it took humans many centuries to recognize the power of negative feedback (Mayr, 1970), it is a design principle widely employed in biological systems. An instance is found in the glycolytic pathway through which cells acquire energy by oxidizing glucose. Glucose is a six-carbon sugar that is split into two molecules of glyceraldehyde-3-phosphate, each of which, as shown in the middle of Figure 2, then undergoes oxidation through the action of *Glyceraldehyde Phosphate Dehydrogenase* (*GAPDH*), with the transfer of a hydrogen ion to NAD^+. The energy liberated in the oxidation is captured in a phosphate bond of 1,3-Diphosphoglycerate before it is transferred in two subsequent reactions to ADP, yielding ATP. Enzymes, whose names are indicated in italics in Figure 2, catalyze each reaction in the pathway. The enzymes constitute the constraints that direct the flow of free energy found in glucose to the synthesis of ATP.

This description of glycolysis treats it as an unregulated reaction in which, when glucose is present, the sequence of reactions ensues until all the glucose has been consumed. This fits the Machamer, Darden, and Craver (2000) characterization of mechanisms as operating from start-up to termination conditions. But such an unregulated process would be very wasteful of glucose, the source of free energy, which may be in short supply. Control is achieved through negative feedback on the second of two reactions at the beginning of the pathway that transfer phosphates from ATP to six-carbon derivatives of glucose. The enzyme that catalyzes this reaction, *Phosphofructosekinase-1*, is an allosteric enzyme, which means it changes its conformation, and hence the reactions it catalyzes, as a result of binding with other molecules. It thus implements time-dependent, flexible control. Even though the enzyme requires ATP to supply a phosphate to Fructose-6-phosphate, the enzyme also reduces the rate of catalysis when ATP is present (note the edge-ended arrow to it on the right which originates in reactions that generate ATP). As a result, glucose is not wasted when ATP is plentiful. Instead, Fructose-6-phosphate accumulates, and as it does so, another feedback loop reduces the rate of the first phosphorylation reaction, keeping glucose from entering the pathway (it is instead transformed into glycogen in a reaction not shown). On the other hand, when ATP is scarce and ADP or AMP, the breakdown products from ATP, is prevalent, *Phosphofructosekinase-1* increases the speed of the reaction, thereby restoring the supply of ATP.

[18] When negative feedback is implemented in devices such as thermostats, there is once again a role for humans in setting the thermostat. The Watt governor does not permit such setting.

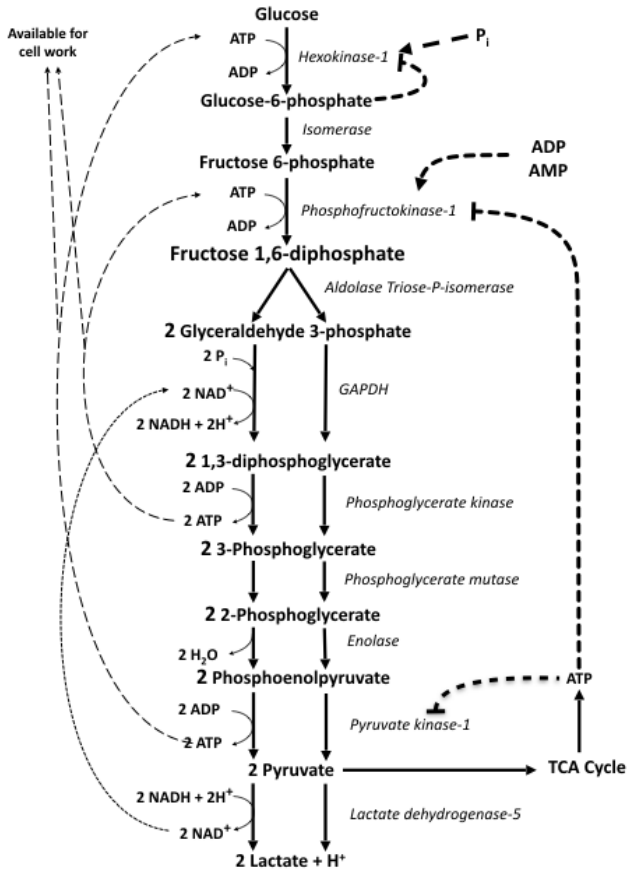
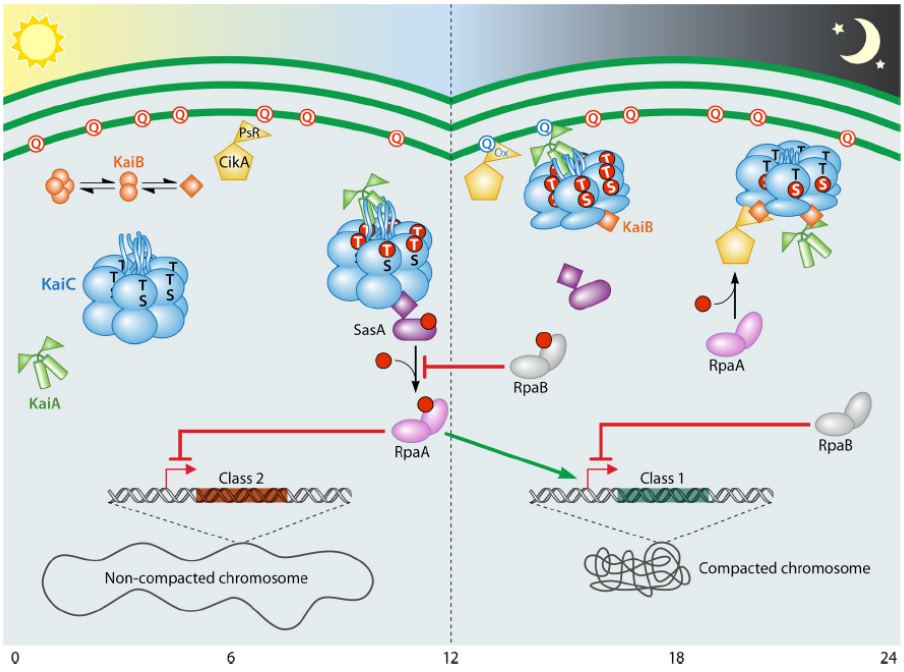


Figure 2. Glycolytic pathway, showing with dashed lines two feedback loops that reduce the consumption of glucose when ATP is plentiful.

This illustrates one use of the design principle of negative feedback to control biological mechanisms. As useful as direct negative feedback is in regulating mechanisms, it is limited in that it is a *quantitative* form of regulation. It cannot direct mechanisms in an organism to function in *qualitatively* different ways at different times—e.g., to switch between metabolizing different sugars or to switch from maintaining cell life to dividing. Bich and Moreno (2016) argue that such regulation (as opposed to maintenance of dynamic stability) requires a regulatory system decoupled from the constraints of the system it regulates. Although the negative feedback mechanisms in the Watt governor and in glycolysis are tightly coupled to the mechanisms they control, they are separate mechanisms employing their own constraints. Constraints in the controlling system could also be responsive to other inputs and so be decoupled from the controlled mechanism, thereby satisfying Bich and Moreno's condition. Rather than explore this as a hypothetical possibility in the examples already considered, though, we will focus on an actual control mechanism that relies on negative feedback in living organisms that is clearly decoupled from the mechanism(s) that it controls—the circadian clock in cyanobacteria.

The circadian clock regulates physiological processes in cyanobacteria by controlling which genes are expressed during which part of the day. As shown in the lower part of Figure 3, two types of promoters bind to the promoter region of genes and initiate their transcription. The promoter regions are thus time-dependent, flexible constraints over gene expression. As a result, genes controlled by Class 1 exhibit peak expression when the clock indicates dusk and those by Class 2 have peak expression when it indicates dawn (circadian time, in which 0/24 corresponds to anticipated dawn, is indicated at the bottom).[19] Individual cyanobacteria track time using a feedback mechanism in which KaiC, shown as a double doughnut to represent the fact that it occurs as a hexamer with two domains, is successively phosphorylated and dephosphorylated through the activities of KaiA and KaiB. When it is completely unphosphorylated (indicating dawn), KaiA, shown in green, binds to the tails extending up from the CII domain of KaiC. This enables phosphorylation to proceed first at the T site and then at the S site. Once it is fully phosphorylated (indicating dusk), a feedback process ensues whereby KaiB binds to the CI domain of KaiC and KaiA detaches from the tails (which then retract) and instead binds to KaiB. This initiates dephosphorylation, first at the T site and then at the S site. When only the T-site of KaiC is phosphorylated, two other molecules, SasA and RpaA, are also phosphorylated, and serve to activate transcription of genes controlled by Class 1 promoters and suppress transcription of genes controlled by Class 2 promoters. At other times, genes with Class 2 promoters are expressed, not those with Class 1 promoters.



[19] Like any representational system, the clock can misrepresent, for example, by indicating dawn when it is midday.

Figure 3. The circadian clock, acting as a controller regulating the expression of genes in cyanobacteria. See text for details. From Cohen and Golden (2015).

This elaborate clock mechanism is decoupled from[20] and operates on the flexible time-dependent constraints—the Class 1 and Class 2 promoters—to regulate gene expression. The importance of decoupling is that it allows the clock mechanism to be operated on by yet another mechanism so that it is entrained to the light-dark cycle in the bacterium's environment. This mechanism consists of two additional proteins, circadian input kinase A (CikA) and Light-dependent period A (LdpA). They sense the redox state of the cell, which serves as proxy for light exposure. We focus just on CikA, which is shown at the top in figure 3. It binds first to quinones and, when it is so bound, to KaiC. When it binds to KaiC, it prevents RpaA from binding. As a result, the circadian control mechanism is capable of altering gene expression in a way that responds to light-dark conditions in the bacterium's environment.

In this section we have illustrated how control of mechanisms within biological systems is performed other mechanisms in those systems. These control mechanisms involve their own set of constraints operating on the time-dependent constraints in the mechanism being controlled. Before extending this discussion of control into the nervous system, there are a couple points to make. First, both in the Watt governor and in the biological mechanisms, the mechanism being controlled exhibits its own behavior as a result of constraints realized in it. The control system does not initiate the behavior of the controlled system but alters its operation along some quantitative or qualitative dimension. Second, while the Watt governor does in fact receive the energy it uses to regulate the valve from the controlled system, that is not typical in human-made machines. In the biological cases, it is an essential requirement. Control mechanisms as well as the controlled mechanisms are parts of the biological system and require energy both to perform work and to maintain themselves as organized, far-from-equilibrium systems. Likewise, while the Watt governor does give the steam engine a (very limited) degree of autonomy, biological systems are much more fully autonomous systems. The control systems involved in glycolysis and those in the circadian clock are components in a broader network of mechanisms that together control each other so as to maintain the biological system as a whole.

6. Climbing a Local Neural Control Hierarchy

The examples of control systems in the previous section allow for a distinction of levels: since a controller, such as a circadian clock, operates on another mechanism, it is appropriately viewed as at a higher level. This process can then be iterated, with mechanisms higher in a local control

[20] It is, in fact, more coupled than Figure 3 indicates, since the synthesis of KaiA, KaiB, and KaiC is also under control of the clock mechanism. Nonetheless, the phosphorylation process is distinct from the synthesis process.

hierarchy operating on lower-level control mechanisms, which then control the basic mechanism. Signaling systems in cells often involve multiple levels of control. We will, however, turn to the nervous system to develop this conception of multiple levels of control. (We refer to the control hierarchies as local because even the highest level in such a hierarchy may ultimately be controlled by some of the mechanisms it controls. This results in a heterarchical network, not a strict hierarchy.) Note here that whereas the mechanist literature has often been concerned with compositional, scalar, and more general causal dependence hierarchies, the notion of a *control* hierarchy (or for that matter a control heterarchy) is conceptually distinct from these and, we contend, more important for understanding biological organization.[21]

The circadian example provides a useful starting point. In mammals, circadian clocks operate in individual cells of the body (using a feedback mechanism involving gene expression, not protein phosphorylation as in cyanobacteria). These clocks each oscillate with slightly different periods and left on their own will not be synchronized with each other. Specialized cells in the suprachiasmatic nucleus (SCN) of the hypothalamus are coupled to each other so that their clocks can synchronize (this involves time-dependent, flexible constraints in the clocks of each neuron that respond to signals from the other neurons). Built out of mechanisms that do a relatively poor job of keeping time, the SCN constitutes a higher-level mechanism that keeps quite accurate time and sends signals to clocks in other tissues of the body. In this case the SCN serves as a high-level controller over the other clocks (operating on constraints in them), which then control gene expression in individual cells of the body. Yet other constraints in SCN cells are themselves operated on by other mechanisms that regulate their basic physiological functioning so that control is ultimately heterarchical.

Although philosophers, focused on cognition, have tended to emphasize high-level information processing activities involved in reasoning, memory, and decision-making and so have focused on processing pathways within the neocortex, it is important to recognize that the brain, including the neocortex, evolved to control other organs of the body. Far from being autonomous from the rest of the nervous system, regions of the neocortex are each highly interconnected with regions in the brain stem and midbrain. Processing proceeds from sensory receptors (including receptors for internal states) via the peripheral nervous system up to higher regions and back down, through the peripheral nervous system, to the muscles. At low levels in the spinal cord, there are ganglia of neurons that integrate sensory signals with motor outputs. Higher levels modulate lower-level activity. This hierarchy culminates in neocortical regions that modulate activity in lower brain regions and ultimately low-level ganglia.

[21] The ambiguous notion of a 'functional level' or a 'functional hierarchy' may have sometimes resulted in these distinctions being blurred, especially that between ordinary causal dependence and control.

Thinking of neocortex as consisting of high-level controllers operating on other brain areas and other organs of the body makes sense evolutionarily. Animals without a neocortex are capable of complex engagements with their environments, including canonical cognitive tasks such as learning, remembering, and decision-making (Ardiel & Rankin, 2010; North & Greenspan, 2007; Stein, Grillner, Selverston & Stuart, 1997). Many of the structures that perform these functions in other animals have orthologs in subcortical regions in mammalian brains. In light of the capacity of the corresponding brain components to carry out information processing tasks needed to regulate behavior in non-mammals, it is plausible to construe subcortical regions in mammals as the immediate controllers of behavior and to treat the neocortex as representing loci of higher-level control within the organism. This perspective reflects that of Jackson (1884), who proposed a hierarchy of control in which motor systems were represented, re-represented, and re-re-represented in subcortical areas, motor areas of cortex, and frontal areas, respectively. Damage to the higher control areas, Jackson argued, freed the lower areas from higher-level control, allowing them to operate independently.

To appreciate this perspective of hierarchical control in the nervous system, we need to set aside the common input-output perspective in which sensory stimuli are received and processed in multiple steps until a decision for action is made, which is then processed through several steps until a behavior is executed. This latter perspective flattens the multiple levels of control. Recently Keijzer (2015), in seeking to understand the evolution of the nervous system, has argued for an alternative to the input-output view, which he terms the *coordination view*. On this view, the first neurons to evolve served to coordinate contractile tissues so that they could work together to move the organism. The first contractile tissues were probably much simpler than contemporary muscles, but we can recognize the need for control if we consider muscles. Among the key components of muscle are actin and myosin molecules. These molecules are so structured that they cyclically construct cross-bridges that allow each to slide along the other. As a result of the constraints imposed in muscle tissue, this process is endogenous and would occur continuously as long as energy is available were it not for the binding of tropomyosin, which blocks the active binding sites between actin and myosin. Tropomyosin functions as a time-dependent switch: when calcium binds to it, it no longer interferes with the binding between actin and myosin. Calcium stores are maintained in the sarcoplasmic reticulum in the cytoplasm from which molecules are only released in response to an electrical current, created by the binding of neurotransmitters to receptors on the muscle cell. The neural signal thus triggers the switch, allowing the endogenous activity of the actin-myosin mechanism to proceed.

The switching mechanism explains how control is exercised over muscle to allow contraction, but control in turn must be exercised over the switch so that contractions occur at appropriate times. Relatively simple model organisms such as the medicinal leech offer suggestive models

as to how such control can be achieved. Leeches can execute a number of motor actions, including swimming, crawling, or feeding, each of which requires activity in a different muscle group. The decision to execute a specific behavior is made separately in each of the 21 segmental ganglia in the nerve cord. These can be exposed so that the activity of neurons that are close to the surface (about a third of the approximately 400 neurons making up each ganglion) can be imaged using a voltage-sensitive dye. Using a neutral stimulus situation which resulted in the leech swimming or crawling equally often, Briggman, Abarbanel, and Kristan (2005) investigated how these neurons exercised control. They first attempted to identify the decision-making neuron by identifying the first neuron to fire in a way that would accurately predict the ultimate behavior. They identified 33 neurons in six preparations that showed a differential response prior to detectable activity in the motor neurons. However, none of these neurons, when either hyperpolarized or depolarized, would bias the decision, indicating that they were not the neurons making the decision. As an alternative strategy, the researchers used principal components analysis and linear discriminant analysis to identify a subpopulation of neurons that as a group showed a differential response before the onset of motor activity. The subpopulation they identified shared only a few neurons with the set of 33 neurons that individually responded early. Among 17 neurons that contributed strongly to the linear discrimination, they found one, 208, that when hyperpolarized biased the leech to swimming and when depolarized biased it towards crawling. These effects depended on neuron 208 being part of the initiating circuit responding to the neutral stimulus—when hyperpolarized or depolarized alone, it did not have these effects.

Although many details remain to be worked out, Briggman et al.'s research revealed that neurons in segmental ganglia operate as a mechanism that controls the leech's various muscles. This provides a model for the multiple neural control systems in more complex organisms. Some of these control basic motor actions, but we, presumably unlike the leech, can make higher-level decisions, such as to go for a walk. This is where higher-level controllers, many located in the neocortex, play a role. As before, all control processes are realized within the biological system. The controllers at the top of a given hierarchy may themselves be controlled by other mechanisms within the biological system. As a result of such a heterarchical control network, control remains in the biological system, rendering it autonomous.

7. Conclusions

We began by identifying a key limitation in applying the mechanist philosophy of science to neuroscience: it has not adequately recognized and factored in the importance of *control* of biological mechanisms. We linked this problem to three other shortcomings of current mechanistic accounts of causation. To address all these problems, we drew upon insights of largely forgotten theoretical biologists such as Pattee and advanced a perspective on causal

production in biological mechanisms that is grounded on the idea that biological mechanisms, like human-made machines, derive their causal efficacy from being *constrained* systems. By restricting some degrees of freedom of its components and thereby enabling the whole mechanism to do things that would otherwise not be possible, constraints determine the causal powers of a machine or mechanism. Of particular importance are those constraints that are flexible and time-dependent. These enable machines to operate in different ways on different occasions.

We then extended this framework by noting that machines perform work, which requires that the constraints direct the flow of free energy in appropriate ways. While constraining the flow of free energy is necessary in any machine, it is particularly important in the case of biological organisms as they are dissipative systems that must secure and manage energy flows within them. Otherwise, they settle into equilibrium with their environment and cease to be alive. We then extended the framework one more time in recognition that biological systems require control in order to cope with a variable environment. Such control involves a controller mechanism operating on time-dependent constraints within a controlled mechanism. While external agents can exercise control over machines, the controllers of biological mechanisms must be other mechanisms within the biological system. Constraints are the theoretical linchpin that ties these capacities together and accounts for biological mechanisms as causally efficacious, far-from-equilibrium dissipative structures that are autonomous in terms of control.

We then applied this framework of controllers operating on time-dependent constraints of mechanisms to several examples, starting with Watt's centrifugal governor for the steam engine. The governor operates by employing a feedback process such that whenever the engine runs too fast, it closes the steam valve and when it runs too slow, it opens the valve. Negative feedback mechanisms are also found widely in biological organisms, and we described the example of feedback control in glycolysis that serves to preserve the supply of glucose by only allowing the glycolytic mechanism to metabolize glucose when ATP is needed.

Feedback control mechanisms are often tightly coupled to the mechanism they are controlling, limiting their ability to qualitatively alter their activity. But as separate mechanisms, they have the potential to be affected by yet other mechanisms acting on time-dependent constraints within them. When this is the case, more complex or higher-order types of control relationships are possible. The circadian clock in cyanobacteria provided an example. It controls the expression of nearly all genes in cyanobacteria and, without influences from elsewhere, maintains a period of approximately 24 hours. But it is also capable of being entrained to the light-dark cycle in the local environment. As a result, the control it exercises varies as conditions change (e.g., with the seasons of the year).

To illustrate how the nervous system allows for a local hierarchy of control mechanisms, we briefly introduced the circadian system in mammals, in which clocks within neurons in the SCN can synchronize with each other and send a regular signal to circadian clocks elsewhere in the organism that then regulate gene expression in each tissue. To appreciate the control function played by neural mechanisms requires modifying the mechanist approach so as to construe neural mechanisms as exercising control over other mechanisms. We illustrated neural control with the decision-making mechanism in segmental ganglia in leeches that controls muscles for swimming, crawling, and feeding. In mammals, this control hierarchy extends to the neocortex in which cortical control systems modulate subcortical ones on down to neurons that operate directly on muscle cells.

The biological control systems we discussed all fit the account of causation in mechanisms we offered. Constraints in both controlled mechanisms and controller mechanisms determine their specific causal powers. Unlike the Watt governor, biological mechanisms are dissipative structures and depend on control mechanisms to direct a flow of free energy to maintain them. These control mechanisms reside within autonomous biological systems and form part of a heterarchical network of controllers. The framework of mechanistic causal production we have offered, unlike that adopted in most mechanistic accounts, captures these important characteristic features of biological mechanisms.

References

Ardiel, E. L., & Rankin, C. H. (2010). An elegant mind: Learning and memory in Caenorhabditis elegans. *Learning & Memory, 17*, 191–201.

Atkins, P. W. (1984). *The second law*. New York: Scientific American Books.

Bechtel, W. (2010). The downs and ups of mechanistic research: Circadian rhythm research as an exemplar. *Erkenntnis, 73*, 313–328.

Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science, 78*, 533–557.

Bechtel, W. (2013). From molecules to networks: Adoption of systems aproaches in circadian rhythm research. In H. Andersen, D. Dieks, W. J. Gonzalez, T. Uebel & G. Wheeler (Eds.), *New challenges to philosophy of science* (Vol. 4, pp. 211–223): Springer Netherlands.

Bechtel, W. (2017). Explicating top-down causation using networks and dynamics. *Philosophy of Science, 84*, 253-274.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 421-441.

Bechtel, W., & Abrahamsen, A. (2011). Complex biological mechanisms: Cyclic, oscillatory, and autonomous. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 257–285). New York: Elsevier.

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.

Bich, L., & Moreno, A. (2016). The role of regulation in the origin and synthetic modelling of minimal cognition. *Biosystems, 148,* 12–21.

Briggman, K. L., Abarbanel, H. D. I., & Kristan, W. B. (2005). Optical imaging of neuronal populations during decision-making. *Science, 307*, 896-901.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Clarendon.

Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge: Cambridge University Press.

Cartwright, N. (2008). Reply to Stathis Psillos. In S. Hartmann, C. Hoefer, & L. Bovens (Eds.), *Nancy Cartwright's philosophy of science* (pp. 195–197). New York: Routledge.

Cartwright, N., & Pemberton, J. (2013). Aristotelian powers: Without them, what would modern science do? In R. Groff & J. Greco (Eds.), *Powers and capacities in philosophy: The new Aristotelianism* (pp. 93–112). New York: Routledge.

Cohen, S. E., & Golden, S. S. (2015). Circadian rhythms in cyanobacteria. *Microbiology and Molecular Biology Reviews, 79*(4), 373–385.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy, 22*, 547-563.

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

Darden, L., & Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Biological and Biomedical Sciences, 33*(1), 1–28.

Esfeld, M. (2009). The modal nature of structures in ontic structural realism. *International Studies in the Philosophy of Science, 23*(2), 179-194.

Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 50–71.

Glennan, S. (2009). Mechanisms. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford handbook of causation* (pp. 315–325). Oxford: Oxford University Press.

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

Hooker, C. A. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science, 18*, 757–780.

Jackson, J. H. (1884). Evolution and dissolution of the nervous system (The Croonian Lectures). *Lancet, 123*, 555–558, 649–652, 739–744.

Juarrero, A. (1999). *Dynamics in action: Intentional behavior as a complex system*. Cambridge, MA: MIT Press.

Keijzer, F. (2015). Moving and sensing without input and output: Early nervous systems and the origins of the animal sensorimotor organization. *Biology & Philosophy, 30*, 311–331.

Kistler, M. (2009). Mechanisms and downward causation. *Philosophical Psychology, 22*(5), 595-609.

Kuhlmann, M. & Glennan, S. (2014). On the relation between quantum mechanical and neo-mechanistic ontologies and explanatory strategies. *European Journal of Philosophy of Science, 4*, 337–359.

Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*(1): 27–39.

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

Maxwell, J. C. (1868). On governors. *Proceedings of the Royal Society of London, 16*, 270–283.

Mayr, O. (1970). *The origins of feedback control*. Cambridge, MA: MIT Press.

McCulloch, W. S. (1945). A heterarchy of values determined by the topology of nervous nets. *Bulletin of Mathematical Biophysics, 7*, 89–93.

Moreno, A., & Mossio, M. (2015). *Biological autonomy: A philosophical and theoretical inquiry*. Dordrecht: Springer.

Nagel, E. (1961). *The structure of science*. New York: Harcourt, Brace.

Nicolis, G., & Prigogine, I. R. (1977). *Self-organization in nonequilibrium systems: From dissipative structures to order through fluctuations*. New York: Wiley.

Nielsen, K. (2010). Representation and dynamics. *Philosophical Psychology, 23*, 759–773.

North, G., & Greenspan, R. J. (2007). *Invertebrate neurobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

O'Malley, M. (2014). *Philosophy of microbiology*. Cambridge: Cambridge University Press.

Pattee, H. H. (1970). The problem of biological hierarchy. In C. H. Waddington (Ed.), *Towards a theoretical biology 3: Drafts* (pp. 117–136). Edinburgh: Edinburgh University Press.

Pattee, H. H. (1971). Physical theories of biological co-ordination. *Quarterly Reviews of Biophysics, 4*(2–3), 255–276.

Pattee, H. H. (1991). Measurement-control heterarchical networks in living systems. *International Journal of General Systems, 18*(3), 213–221.

Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia.

Ross, D., Ladyman, J., & Spurrett, D. (2007). In defence of scientism. In J. Ladyman & D. Ross, *Every thing must go: Metaphysics naturalized* (pp. 1–65). Oxford: Oxford University Press.

Ruiz-Mirazo, K., & Moreno, A. (2004). Basic autonomy as a fundamental step in the synthesis of life. *Artificial Life, 10*, 235–259.

Ruiz-Mirazo, K., Peretó, J., & Moreno, A. (2004). A universal definition of life: Autonomy and open-ended evolution. *Origins of Life and Evolution of the Biosphere, 34*, 323–346.

Sklar, L. (2013). *Philosophy and the foundations of dynamics*. Cambridge: Cambridge University Press.

Smart, J. J. C. (1963). *Philosophy and scientific realism*. London: Routledge & Kegan Paul.

Stein, P. S. G., Grillner, S., Selverston, A. I., & Stuart, D. G. (Eds.). (1997). *Neurons, networks, and motor behavior*. Cambridge, MA: MIT Press.

Turvey, M. T. (1977). Preliminaries to a theory of action with reference to vision. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing* (pp. 211–265). Hillsdale, NJ: Erlbaum.

van Gelder, T., & Port, R. F. (1995). It's about time: An overview of the dynamical approach to cognition. In R. F. Port & T. van Gelder (Eds.), *Mind as motion* (pp. 1–43). Cambridge, MA: MIT Press.

Varela, F. J. (1979). *Principles of biological autonomy.* New York: North Holland.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2008). Response to Strevens. *Philosophy and Phenomenal Research, 77*(1), 193–212.

Yates, F. E. (1979). Physical biology: A basis for modeling living systems. *Journal of Cybernetics and Information Science, 2*, 57–70.