Biological Theory
https://doi.org/10.1007/s13752-025-00516-5

THEMATIC ISSUE ARTICLE

Open-Ended Control Versus Closed-Ended Control: Limits of Mechanistic Explanation

Jason Winning[1]

Received: 3 March 2025 / Accepted: 6 October 2025

Abstract
Some recent discussions of mechanistic explanation have focused on control operations. But control is often associated with teleological or normative sounding concepts like goals and setpoints, prompting the question: Does an explanation that refers to parts or entities within mechanisms "controlling" each other thereby fail to be mechanistic? In this article, I introduce a distinction between open-ended and closed-ended control. I then argue that explanations that enlist control operations to do explanatory work can count as mechanistic in the New Mechanist sense only if such control operations are closed-ended, not open-ended. In certain scientific fields that incorporate control operations within their mechanistic models, for example, systems neuroscience, the state of the science is such that this requirement often cannot be met. We should therefore distinguish between models/explanations that are "mechanistic" in the strict (New Mechanist) sense and those that are "mechanistic" in a weaker sense employed by fields in the latter category.

Keywords Control · Mechanisms · Mechanistic explanation · Teleological explanation · Teleology

Introduction

> *One of the beautiful things about mechanisms is, they don't think. They'll either function or malfunction.*
>
> —Brad Bachelder, firearms expert, Forensic Files episode "Murder, She Wrote".

Proponents of mechanistic explanation have long counted amongst its virtues that it avoids the need to invoke a designer, goals, final causes, or other normative or mentalistic-sounding notions. Mechanisms operate the way they do at a given moment solely because of the constituents they possess at that moment and how they are organized. According to the New Mechanist philosophy of science, mechanistic explanations also refer to the "operations" or "activities" of the components and treat them as similarly describable in non-intentional, non-teleological terms.[1]

Recently, philosophers (for example, Bechtel and Bich 2021) have called attention to the fact that mechanistic explanations in biology often include *control*: some of the operations of mechanistic components are control operations, and some mechanisms operate on others by controlling them. But control is often associated with teleological or normative concepts, such as that of a goal, setpoint, or "normal range" of operation that the controller acts to achieve or maintain. Sometimes "control" is even defined explicitly in terms of paradigmatically intentional states like desires.[2] This is in contrast to other "stock in trade" (as Kauffman (1971) puts it) mechanistic operations like transport or ligand-induced conformational change, which are generally not thought of as being inherently normative or teleological. Can an explanation that refers to parts or

✉ Jason Winning
jason.winning@gmail.com

[1] Madison, WI, USA

[1] Some New Mechanists prefer to use the word "entity" instead of "component" or "part," and "activity" or "interaction" instead of "operation." For purposes of this article, I regard these terms as interchangeable. For discussion of this terminology, see Glennan (2017, pp. 19–22).

[2] For example: "A *control system* may be defined as a collection of interconnected components that can be made to achieve a desired response in the face of external disturbances" (Khoo 2018, p. 1).

Published online: 27 October 2025

mechanisms "controlling" each other really be a *mechanistic* explanation?

My conclusion in this article will be that whether or not control can factor into a mechanistic explanation, strictly so-called, depends on whether the control operation is understood in an *open-ended* or *closed-ended* way. An open-ended controller is one that does not operate with a fixed input–output relation but can be relied upon to select whatever the appropriate way to intervene on a target process is *because of* the appropriateness of that way of intervening. Mechanistic explanations can only include closed-ended control; the enlistment of open-ended control operations to do explanatory work renders an explanatory model at least partly teleological (and therefore at least partly non-mechanistic) in nature.

While this is true of mechanistic explanations as philosophers (including New Mechanists) use the term, there is a much more impoverished sense of the word "mechanism" that is employed in certain fields, such as systems neuroscience, where mechanistic explanation in the strict sense is often not possible yet. The distinction between open-ended and closed-ended control I will introduce in this article can also be used to clarify differences in usage of "mechanistic explanation" between fields.

In what follows, I first introduce some reasons for thinking that it is important to make room for control operations in accounts of mechanistic explanation in biology. Next, I consider several arguments against inclusion of control in mechanistic explanations and show why they fail. I then demonstrate the difference between understanding control operations in an open-ended versus a closed-ended way using an example and explain why mechanistic explanations strictly so-called should not include open-ended control operations. Finally, I consider how open-ended control is discussed in some fields, including systems neuroscience where it commonly figures into "mechanistic" models/explanations, and therefore argue that we should distinguish between strict and weak ways of understanding mechanistic explanation.

Mechanistic Explanations and Control

"New Mechanist" philosophers of science such as Bechtel (Bechtel and Richardson 1993; Bechtel and Abrahamsen 2005), Glennan 1996, 2017; Machamer et al. (2000) describe mechanistic explanations as explaining a system's production of a phenomenon of interest by reference to the causal operations that the system's parts perform (and that are performed on them), as well as the way these parts and operations are organized into a whole. This view of explanation was based on an analogy between biological mechanisms and human-built machines: "By calling the explanations *mechanistic*, we are highlighting the fact that they treat the systems as producing a certain behavior in a manner analogous to that of machines developed through human technology" (Bechtel and Richardson 1993, p. 17). While there is truth to this analogy, it can suggest a misleading view of biological mechanisms as mere *production* mechanisms whose construction and ongoing maintenance are taken for granted (Winning and Bechtel 2018). The internal operations of a human-built machine, an electric fan for example, are usually totally separate from the operations that went into initially constructing it or that go into fixing it if it needs repair. Its operation is "ballistic" in the sense that it carries out a predictable sequence of internal causal processes that are constrained merely by the device's own static internal structure; there is no need for internal *control*. Control of the system is usually done externally by a human operator; the device's internal operation can be accounted for in a mechanistic way without reference to the notion of control.

What exactly does "control" refer to, in the context of mechanistic explanation? To see this, consider the fact that biological systems are very much unlike typical human-built or human-designed machines in at least three ways:

- Whereas human-built machines resist thermodynamic equilibrium forces primarily due to the static rigidity of the materials out of which they are composed (metal alloys, plastics, and so on), organic systems are highly dynamic and maintenance of their structures far from equilibrium requires constantly varied work that must be internally controlled.
- Whereas human-designed machines are usually built and repaired by external human agents, organic systems must continually construct and reconstruct their own structures, repair their own structures, and modify their own structures adaptively to maintain internal functioning in changing environmental circumstances.
- Whereas most functions of human-built machines are designed to function "ballistically" in the sense described above (with certain exceptions like guided missiles), internal control (especially feedback control) frequently occurs as one or more steps of an organic system's internal production processes.

"Control," then, refers to operations performed by mechanisms or their parts/entities that construct, repair, modify, or maintain other processes, mechanisms, and/or parts of mechanisms in ways that are systematically sensitive to the surrounding and/or internal conditions of the system they are part of.[3]

[3] The "target" processes or mechanisms that controllers act upon are sometimes referred to as "production" processes or mechanisms

The features of biological systems in the above list are only possible due to the ubiquitous presence of control systems and control operations in biological mechanisms (Pattee 1973). But taking their role into account in New Mechanism is not a trivial matter of appending "control" to the list of operations a part of a mechanism might perform. The presence of internal control means that non-reducible, holistic *organization* takes on an even greater importance in mechanistic explanations, since the presence of feedback control loops introduces nonlinear phenomena that are difficult to explain or predict in terms of a static, sequential arrangement of parts and operations (Bechtel and Abrahamsen 2011). Further, control operations are, themselves, qualitatively distinct from other mechanistic operations and inherently more complex, for example, because they often have to integrate information from multiple sources and use this information to intervene on other processes or mechanisms in subtle ways.

Given these facts, it might be wondered whether a new version of New Mechanism that gives a central role to control represents a modification to, or a departure from, mechanistic explanation properly so-called. In the following sections, I consider and evaluate several arguments against the inclusion of control in mechanistic explanations.

The "Black Box" Argument

One way to argue that control does not belong in mechanistic explanations derives from the inherent complexity of control operations. Components of mechanisms that perform control operations (that is, controllers) have multiple parts that play various functional roles. When these are fully spelled out in the mechanistic model, they will be sufficient to explain what the controller does. Including both the controller itself as well as its subcomponents would then be redundant, and perhaps even a problematic form of overdetermination. To prevent this, the argument might run, mechanistic explanations should only include the subcomponents and their more basic operations, so that the controller and its control operations no longer feature in the model.

Of course, one can always get around this problem by omitting the mechanistic details of the controller and treating it as a single "black box"—a part within a larger mechanism that performs a control operation within that mechanism. However, it might be argued that a mechanistic explanation that does not detail the internal complexity of the black box would therefore be incomplete, or that this would prevent the explanation from being thoroughly mechanistic in nature. One might argue that control is what Craver refers to as a *filler term*, a term that is "used to indicate a kind of activity in a mechanism without providing any detail about how that activity is carried out" (2006, p. 360). According to Craver,

> filler terms are barriers to progress when they veil failures of understanding. If the term "encode" is used to stand for "some-process-we-know-not-what," and if the provisional status of that term is forgotten, then one has only an illusion of understanding. (Craver 2006, p. 360)

If the underlying mechanistic details of the controller are fully spelled out, then these details can simply replace the filler term "control" in the explanation, so that it will no longer occur. To avoid having a mere mechanism "sketch" or mechanism "schema," and instead have a complete mechanistic explanation, the argument might run, one must therefore eliminate the control operations and replace them with more fully spelled out mechanistic details.

However, Glennan points out that "complete" is a relative term when applied to mechanistic explanations. A mechanistic explanation need only be "complete at a single level of the mechanism" to avoid being a mere sketch or schema (Glennan 2017, p. 76; cf. Craver and Kaplan 2020). As long as the explanation details the relevant operations at a given compositional level below where the phenomenon of interest occurs, such an explanation can be complete, even if the operations it specifies could themselves potentially be fleshed out further. A mechanistic explanation does not need to plumb the depths of every compositional level all the way down to elementary fields and particles. If control operations occur at a compositional level below the level of the whole that manifests a phenomenon of interest, the fact that such operations are themselves realized by sub-mechanisms does not mean that such operations cannot feature in a complete mechanistic explanation.

What Kinds of Operations Can Mechanistic Explanations Include?

But one might still argue that because control operations are more sophisticated than, and qualitatively distinct from, other kinds of mechanistic operations, they should not be included in a properly complete mechanistic explanation. Kauffman argued that operations appealed to in mechanistic operations should be "simple in the sense that it is by articulating together more than one of these well understood

when they do not themselves play (or are not being considered to play) a control role; see Bechtel and Bich (2021, pp. 1–2). This is not intended as a definition; ways of drawing the line between control and non-control operations are considered in the section "Mechanistic Accounts of Signal and Control Pathways: Closed-Ended Control".

processes that we seek to explain more complex processes" (1971, p. 268). Correspondingly, it might be argued that referring to control operations does not sufficiently amount to explaining what is complex in terms of what is simple.

This raises the question of what kinds of operations are out-of-bounds in mechanistic explanations. While New Mechanist philosophers of science have sometimes listed examples of operations that might be included in a mechanistic explanation (for example, "biosynthesis, transport, depolarization, insertion, storage, recycling, priming, diffusion, and modulation" (Machamer et al. 2000, p. 8)), they have generally avoided providing clear criteria for what can and cannot count as a part or operation included within an explanation without rendering the explanation non-mechanistic. One notable exception is the following passage:

> Though at times I adopt a specifically causal-mechanical view of explanation (see Craver 2007), and so will describe the ontic structures involved in explanation as causal or mechanistic, I intend the term *ontic structure* to be understood much more broadly. Other forms of ontic structure might include attractors, final causes, laws, norms, reasons, statistical relevance relations, symmetries, and transmissions of marks, to name a few. (Craver 2014, p. 29)

This suggests the following criterion: if an explanation relies on any of the things in Craver's list (attractors, final causes, laws, norms, reasons, statistical relevance relations, symmetries, or transmissions of marks) to carry explanatory heft, then that explanation is excluded from being entirely mechanistic. At best, it might be a hybrid of mechanistic and some other kind of explanation.[4]

Important inclusions on Craver's list are final causes, norms, and reasons. More generally, explanations that rely on unreduced teleological or intentional properties to carry explanatory heft cannot be mechanistic. For example, suppose I am developing an alternative explanation of how the phenomenon of respiration occurs. I note that fumarase is clearly a key component playing a role in this process. On this basis, I decide that fumarase's presence must help to explain respiration because fumarase naturally does whatever would best facilitate respiration. Of course, this operation typically manifests in its reacting with fumaric acid, but I infer that if it could react with some other molecule that would be more helpful, it would do that instead. So, in my explanation of respiration, I characterize the operation of fumarase as that it "does whatever is the best thing within its power to facilitate respiration."

The problem here, of course, is that something like "do whatever is best" cannot count as a mechanistic operation (at least, not one that can do explanatory work in such an explanation), because it includes reference to what is "best," a normative concept. Similarly, we could not characterize fumarase's operation as being to "help secure the goal of respiration," because this relies on the end state to help explain the occurrence of respiration, whereas mechanistic operations must refer only to states of the mechanism that occur during the production process, not after it has completed.

Suppose we instead say that the fumarase has the *current ability* (during the production process) to do whatever helps to secure the goal of respiration. This would at least avoid the appearance of backward causation. The problem that remains is that the component itself is now being described in a way that implies possession of a property verging on an open-ended general adaptiveness or instrumental rationality or practical wisdom, such that it can be relied on to select the *appropriate* means of getting the job done, *because* it is the appropriate means. That a component can select a means *because* it is appropriate (rather than selecting the means that accidentally happens to be the appropriate one) is a form of guidance by goals or intentional states that cannot do explanatory work in a mechanistic explanation unless that explanation includes mechanistic details explaining how such guidance comes about. This is why Dennett (1973) argued that when behaviors are looked at from a purely mechanistic standpoint, they are necessarily treated as "tropistic," or explainable without reference to appropriateness with respect to a goal.

It is important to note here that mechanistic explanations can include parts that have intentional states and the ability to select actions on the basis that they are appropriate for a goal, so long as these intentional states and abilities are not doing explanatory work in such explanations. For example, one can mechanistically explain how one group of people won a game of tug-of-war against another group of people by setting aside their goals and mental states and appealing solely to the physical forces exerted by the members of each group on the rope.

Mechanistic Accounts of Signal and Control Pathways: Closed-Ended Control

As philosophers like Bechtel and Bich (2021) emphasize, mechanistic explanations in biology often detail pathways of control. A control pathway involves detection of some condition, which triggers a signal pathway, which finally culminates in a change in gene expression or the operation of some particular mechanism that is the *target* of control. A complete mechanistic description of a control pathway

[4] We might also add that mechanistic explanations should not depend on things like historical properties or what Shoemaker (1980) calls "mere-Cambridge properties."

should explain what kinds of chemical reactions correspond to "detection" of the condition in question, the sequence of chemical reactions corresponding to the signal pathway, and the systematic relationship between detected conditions and the final control outcomes (for example, changes in gene expression or change of the operation of some target mechanism), referred to as the "input–output relation" (Shinar et al. 2007). In order for this description to be mechanistic, the functional outcome of the signal pathway should be explained as a causal result of the signal pathway, and initial detection and its generation of the signal pathway should also be explained in causal terms, omitting unreduced teleological terms like "appropriate," "too many," or "enough."

Given these restrictions, it might seem that no resources are left over to differentiate control from other mechanistic operations. Any mechanistic operation can be characterized as a "detection" in a certain sense, since it will be a response to the presence of some condition or other. Any causal effect of such a "detection" can then be described as a "signal," since it will carry information (in the sense of causal covariation) about the present condition that was detected. Any downstream causal effects of the "signal" might then be counted as "control" operations.[5]

However, Bechtel and Bich (2021) argue that this would be to lose sight of the important difference between production and control. One way to distinguish control from non-control operations in mechanistic terms was suggested by MacKay, who wrote that in control systems,

> the input, A, determines the form of the output, B, without supplying all the energy of B … the energy of A is at least partly devoted to altering the structure through which the energy for B is channeled—altering the coupling between the output, B, and its internal energy supply (MacKay 1964, p. 311).

The idea is that only if the signal pathway has a major qualitative effect on the mechanisms affected, while supplying a disproportionately low amount of the energy and/or resources necessary for that functional outcome, is it playing a *control* role.

An additional condition is necessary, however. Again, there must be a systematic relationship between the condition detected at the outset of the control process and the resulting qualitative effect on the controlled mechanism(s) (the input–output relation). A negative feedback control system, for example, has a qualitative effect that tends to affect the variable being measured so that it moves that variable closer to a particular value (the setpoint). Many other systematic relationships are possible in control systems as well.[6]

Pattee (1971, 1972) argued further that part of the essence of a control system is that the same *type* of condition can be responded to in the same *type* of way multiple times—in other words, that exact repetition is part of the essence of control. However, since a real-world dynamical system never behaves in exactly the same way on different occasions, exact repetition can only happen if the input–output relation is defined in terms of *ranges* in the state space rather than *points*. Controllers in physical systems must be sensitive to the ranges that input values fall into, rather than the exact values of those inputs. Similarly with outputs: the controller then produces an output behavior that falls anywhere within a certain range, without determining exactly where in the range it will fall. Pattee argues that it is exactly this flexibility of control systems that is necessary for the ability of biological systems to remain stable under perturbations and to resist equilibrium forces.

Importantly, though teleological considerations cannot enter into whether an operation in a mechanistic explanation counts as a control operation and cannot contribute to the explanatory role of the operation, control operations and their functional outcomes can still be associated with teleological roles. For example, researchers might ask "How is the signal that the cell should divide generated?" and yet be asking for a mechanistic explanation for the release of a certain signaling molecule, even though "should" occurs in the question. A researcher might regard the signaling molecule as "teleological" in the sense of having the "purpose" of

[5] It is important to emphasize, as I have done in the preceding paragraph, that the notions of "detection," "signal," "input," and "output" are being used in the very broad sense typical in biology, that could be used in reference to almost any causal process. What I mean by "detection" is captured by Dretske's (1981) and Cummins and Poirier's (2004) notion of *indication*; A "indicates" B if and only if it carries information (or carries a "signal") about B in the sense of causal covariation. An "input" X of Y here refers to some aspect or site X of Y that can be causally affected by something else, or some effect X that something else can have on Y. An "output" X of Y here refers to some aspect or site X of Y at which or by which Y can causally affect something else, or some effect X that Y can have on something else.

[6] Negative feedback is a very common control system paradigm but there are other forms of control as well, and while negative feedback often implies comparison to an internally represented setpoint, this is not necessarily the case. Milsum (1966, p. 11; cf. Willems 1995) distinguishes between "active control" and "passive control." With active control, there is an explicit feedback signal that is compared with an internal reference, and the "error" or difference between them is used to generate a control output. Passive control does not incorporate actuation driven directly by a mismatch between feedback measurements and an internal reference and instead relies on the controller's natural tendency to respond to conditions in systematic ways without comparison to an internal representation being necessary. Another important distinction is between closed-loop and open-loop control (Milsum 1966, p. 44), which should not be confused with the closed-ended versus open-ended control distinction introduced later in this section.

triggering cell division, but the characterization of it as having a purpose (rather than merely a causal effect) will carry no weight in the explanation. Such a mechanistic explanation does not actually require that the signal have, as semantic content, anything about what "should" happen.

Similarly, the setpoint of a negative feedback control system might be characterized by a researcher as the "goal" of the system, but this attribution of a goal does not mean that the explanation of how the system works is thereby teleological in nature. A mechanistic explanation can treat the signal's tendency to bring the system towards the setpoint merely as a causal disposition. The idea that the measured variable *should* be moved toward the setpoint will not then play any role in explaining *how it does* move toward that point.

Given the above considerations, we might define two ways that control operations can be invoked in a mechanistic explanation: closed-ended control versus open-ended control. Closed-ended control operates in a well-understood, predictable way, with a fully specified input–output relation. The input and output conditions are fully specified as a closed set of possibilities, as is the systematic relation that maps inputs to outputs. Open-ended control on the other hand, which is usually what is meant when discussing whether a human being is "in control" of something, means that the controller has a capacity to exert influence in a way that tracks *appropriateness* across an open-ended range of kinds of circumstances (in the language of philosophy of action, such control is "reasons-responsive"; McKenna 2017) and defies characterization in terms of a well-defined input–output relation.

When Does Incorporation of Control Render a Model Non-Mechanistic?

Reliance on control operations can prevent an explanatory model from being mechanistic when those operations are understood as bestowing onto the controller an *open-ended* capacity to anticipate and execute whatever kind of modulation will be advantageous to the system. Control operations understood in this kind of way run afoul of Dennett's requirement that mechanistic operations must be "tropistic," referenced above. I now demonstrate what this would look like with an example.

Weber's Law is a systematic relationship that is observed in a wide range of kinds of visual systems of many species as they adapt to varying light levels without a corresponding decrease in resolution:

> As the background light level increases, the sensitivity of the visual system is decreased, which allows for operation over a huge range of light levels. From a dim starlit night to a bright sunny day, the background light level varies over 10 orders of magnitude (Hood and Finkelstein 1986), and yet our eyes continue to operate across all these levels without becoming saturated with light. The visual system accomplishes this by ensuring that its sensitivity varies approximately inversely with the background light, a relationship known as Weber's law (Keener and Sneyd 2009, p. 893).

This relationship is also observed in bacterial chemotaxis: "chemotactic cells … display 'logarithmic' tracking or sensing, characterized by a constant amplitude response when moving in a gradient that increases exponentially or nearly exponentially" (Sourjik and Wingreen 2012, p. 264). Suppose researchers at (fictional) lab X want to incorporate this characteristic into their mechanistic model of chemotaxis in *E. coli*. They identify an interaction network module of proteins that appears to be the location where control of the dynamic range of discrimination based on the ambient level of ligand concentration likely occurs, and include this module in their model, though they are unable to determine the underlying details by which this control is executed. Suppose further that researchers from lab Y object to this mechanistic model on the grounds that merely including a "Weber's Law module" controlling the dynamic range of discrimination is unilluminating; without detailing the underlying mechanism of that module, it adds nothing to our understanding of how chemotaxis works in *E. coli*.

The researchers of lab X respond by arguing that it is well known that Weber's Law represents an evolutionary optimization; we can characterize the module's operation as maintaining an optimal level of sensitivity to the changes in ligand concentration corresponding to the ambient level. Given an ambient level detection signal, we can predict the effect of the resulting sensitivity adjustment on flagellum behavior based on what would be optimal or most adaptive for the organism. The lab X researchers argue that these predictions are more robust and well-grounded when attributed to a module that is included as a part of the mechanism, rather than to a mere inductive generalization from observed behaviors.

Of course, the researchers from lab Y would not find this argument satisfactory, and it is not an argument biologists would actually make. The problem is that the control operation itself is being defined or characterized in terms of what would be *optimal* or *most adaptive* for the organism, whereas, of course, a bacterium could not possibly have a control mechanism with information processing capabilities sufficient to entertain questions about what is optimal or most adaptive for it. Optimality models are a perfectly

valid way of reasoning about biological systems, but the explanations they provide are teleological, not mechanistic, in nature.

We might characterize the operation of the module by saying that it "optimizes" the dynamic range of discrimination. Or we might characterize it by saying that it adjusts the dynamic range in a way that tends to correspond with Weber's Law. The first of these characterizes the control operation in teleological terms; the other characterizes it as a non-teleological, causal disposition. On the second characterization, the fact that the dynamic range of discrimination of changes in ligand concentration ends up being fairly close to what would be optimal for the *E. coli* is an accidental feature of the module itself, not an outcome it is part of the module's intrinsic feature set to produce.

While the second characterization does not provide mechanistic details of the module's operation, inclusion of the module should not be considered off-limits and can still potentially add to the explanatory power of the mechanistic model in several ways: firstly, by localizing this functional role within the mechanistic model; and secondly, by placing the operation within a sequential ordering of operations within the mechanism. Both of these can allow for predictions about how the mechanism would respond to internal changes and for possibilities of experimental manipulation that mere knowledge that the dynamic range adjustment is observed to occur does not.

If a component in a biological system is found to play a control role that allows novel predictions, this is not necessarily because that component embodies an ability simply to do what would be optimal *because* it is optimal, or to modify some process in an adaptive way *because* it is adaptive. Control systems do not generally have anything approaching practical wisdom, allowing them to just "know" the objectively best way to modulate whatever process they are modulating in given circumstances; no automatic tendency towards absolute optimality or absolute adaptiveness is implied by the fact that a component plays a control role in a system. Only if something like this were implied by control would control be too teleological to count as a mechanistic operation. The upshot is that as long as control operations are fully defined in a closed-ended way with a determinate input–output relation, their incorporation in an explanatory model is no obstacle to that model's being mechanistic.[7]

Open-Ended Control: Applications

Up to now, this article has been concerned with the question of whether control operations belong in mechanistic explanations. In order to answer that question, I introduced the distinction between closed-ended and open-ended control. This might seem like a contrived distinction, having no relevance to real-world science, now that modern biochemistry and neuroscience have made earlier vitalist modes of explanation obsolete. Does the notion of open-ended control have any application in modern-day biological theories or explanations? And if so, would any modern-day investigator or theorist actually invoke open-ended control operations in an explanation that is intended to be mechanistic?

In this section, I examine several references to open-ended control in biology and philosophy of biology literature as they occur in connection to purportedly mechanistic as well as non-mechanistic explanation. While no one, as far as I know, has explicitly characterized a mechanistic operation as one involving open-ended control (because as far as I know, the closed- versus open-ended control distinction has not been made explicit before now), we may use the criteria I have laid out in previous sections to categorize an author's description of a control operation as implicitly either open-ended or closed-ended.

Walsh et al.: "Agential Dynamics"

Walsh and colleagues employ their conception of *agency* in an attempt to "bridge explanatory gaps left by conventional approaches" (Sultan et al. 2022, p. 1) to understanding evolution in biology. They argue that to understand evolution, you must view organisms as *agential systems*, systems in which the "components are sensitive to the shifting contexts provided by the goal-directed activities of the entire system" (2022, p. 8). They continue: "The system-to-component explanation afforded by the agential perspective explains why, in any given context, the components have the properties they have. In agential systems, this explanation cannot be furnished from the mechanism perspective" (2022, p. 8). In other words, the operations of the components are dependent on the current goals and goal-directed activity of the system.

They argue that in order to understand how novel complex traits come into being, you must understand the functioning of gene regulatory networks, which shape the

[7] While an explanation's being mechanistic requires its control operations to be closed-ended, this does not mean that closed-ended control cannot occur in other types of explanations. Closed-ended control might be cited within a larger non-mechanistic explanation, for example a teleological or intentional explanation. For example, a typical intentional explanation might explain my going to the kitchen to get a beer by reference to my desire for a beer and my belief that there is a beer in the fridge. One might add to this explanation an account of activation of the median preoptic nucleus (sometimes said to register fluid balance; McKinley et al. 2019) in the hypothalamus, described as a closed-ended control mechanism, to help explain the presence of the desire. In this case, perhaps it may be debated whether the total explanation still counts fully as an "intentional" explanation, but if not, we may perhaps count it as a hybrid explanation containing mechanistic and intentional elements.

developmental and functional processes in ways that are appropriate to the organism's goals, and that are even "creative" (Sultan et al. 2022, p. 7). The capacity for adaptive novelty means that these networks cannot be understood by reference to the parts, operations, and mechanistic organization of the system alone. Instead, the operations of these systems must be described in terms of "agential dynamics," that is, a capacity to act in the ways that best serve the system's goals, even in novel conditions: "a growing body of work suggests that even when encountering novel conditions, developmental systems may generally be biased toward producing functionally integrated, adaptive phenotypes" (2022, p. 6).

In this way, "agency underwrites a distinctive mode of explanation; because an agent is capable of attaining and maintaining stable endpoints that reliably secure its stability, one can cite the stable endpoint to which the system tends in explaining its activities" (Sultan et al. 2022, p. 5). For example, writes Walsh, "genes collectively have it in their repertoires to produce the appropriate output, under previously unexperienced circumstances" (Walsh 2015, p. 125). This bears some similarity to the earlier fictional characterization of the operation of fumarase as that it "does whatever is the best thing within its power to facilitate respiration": an operation is being defined, not in terms of a determinate input–output relation, but as reliably manifesting whatever the appropriate way to intervene on a target process is *because of* the appropriateness of that intervention. This is exactly how I defined open-ended control earlier. Walsh and his collaborators are essentially arguing that, due to the ability to manifest novel and creative solutions to challenges, developmental regulatory systems must be described in terms of what I am calling open-ended control and then, inferring from this, that we cannot rely on mechanistic explanation to understand them.

Walsh and his colleagues are explicit about the fact that they consider a mechanistic explanation to be insufficient to capture control operations that are best understood in terms of a goal-directed bias, rather than closed-ended input–output relations. However, as we will see in the following subsection, other authors claim to offer *mechanistic* models that incorporate control operations that are described as implicitly open-ended.

Braver et al.: "Mechanisms of Motivation–Cognition Interaction"

Braver et al.'s goal is to review progress that has been made towards providing "an account of motivation–cognition interaction in terms of the neural mechanisms that enable such interactions to occur" (Braver et al. 2014, p. 453). One of the neural mechanisms that they highlight involves the anterior cingulate cortex (ACC), which they claim "might serve as a critical interface between motivation and executive function, by computing the 'expected value of control'" (Braver et al. 2014, p. 457). By "expected value of control" they mean that the ACC computes the cost/benefit ratio of exerting certain kinds of cognitive control, in order to help decide what kind should be exerted at a given moment (Shenhav et al. 2013).

To do this, the ACC weights future rewards that might result from exerting the type of control in question versus more immediate rewards, as well as the cost of exertion of control in terms of the strength of the control signal necessary and loss of the benefit of other competing control signals. While their model separates these out as variables and specifies that the ACC generates the control signal that maximizes the cost/benefit ratio, the model does not spell out how this maximization function is implemented, nor how the candidate control signals are enumerated, or how reward value is assigned. Instead, these functions are often described by the authors in teleological terms such as "appropriateness." They write that "activity in dACC was required to specify the identity of the task-appropriate control signal" (Shenhav et al. 2013, p. 226) and that "representations of affective/motivational significance [are] conveyed to the dACC in order to appropriately modify processing to influence autonomic states as well as changes in overt behavior, including emotional expressions" (2013, p. 231), and so on.

While the authors claim to offer an "account that is mechanistically explicit" (Shenhav et al. 2013, p. 222), according to the argument I have developed thus far, they have not yet offered a thoroughgoingly *mechanistic* explanation for how modulation of cognitive control occurs in the cases they are interested in (such as the Stroop Task), since they are not able to fully explicate the exact input–output relation of the ACC, that is, to specify the range of possible inputs and outputs and the mapping between them as a form of closed-ended control. It may, however, be correct to say they have identified some mechanistic parts and operations and a hypothesis about the functional role of ACC that can point in the direction towards a future mechanistic explanation.

Braver et al.'s (2014) review also summarizes the "dual mechanisms of control" (DMC) framework, according to which the variability of cognitive control results from the interplay of distinct modes of control, proactive and reactive: "Proactive control reflects the sustained and anticipatory maintenance of goal-relevant information within lateral prefrontal cortex (PFC) to enable optimal cognitive performance, whereas reactive control reflects transient, stimulus-driven goal reactivation that recruits lateral PFC (plus a wider brain network) based on interference demands or episodic associations" (Braver 2012, p. 106). Each of these

modes has distinct costs and benefits. Braver argues that the proactive mode is more effective overall because the cognitive system will be better prepared and configured for the task via activation of task goals, plans, and other representations. However, this mode is also more costly overall since it requires active, ongoing maintenance of goal/task representations in the lateral PFC and greater working memory utilization.

According to Braver et al., "the dual mechanisms of control (DMC) framework suggests a specific mechanism" by means of which cognitive activity in the PFC is modulated so that one or the other mode of control is activated (Braver et al. 2014, p. 457). Key to this mechanism is the midbrain dopamine system, which is "postulated to regulate the contents of PFC via a *dynamic updating* mechanism sensitive to reinforcement contingencies" (Braver et al. 2007, p. 78). Again, operation of this mechanism is characterized in terms of *appropriateness*: "Computationally, proactive control is thought to be achieved via dopaminergic inputs to lateral PFC, which enable both appropriate goal updating (via phasic dopamine signals) and stable maintenance (via tonic dopamine release) in accordance with current reward estimates" (Braver et al. 2014, p. 457). More specifically, the dopaminergic system

> is postulated to … play a critical role in learning based on predictions of expected reward (i.e., reinforcement-based learning; Schultz, Dayan, and Montague 1997). Because of this learning role, the [dopaminergic] system can self-organize to develop the appropriate timing of gating signals to enable the appropriate updating and maintenance of relevant context. As such, the system is not a "homunculus," in that it uses simple principles of learning to dynamically configure and adaptively regulate its own behavior. (Braver et al. 2007, p. 79)

While Braver and colleagues do explain in other publications how the dopaminergic system can facilitate basic reinforcement learning in PFC, not enough is explicated (or known) about how other inputs to the system supply context variables and how they are taken into account, nor how tonic dopamine signals facilitate maintenance of goal representations in order for the authors to have described the dopaminergic system in their model in terms of closed-ended control. This is true even if they have said enough to prevent the dopaminergic system from having to rise to the level of a "homunculus"; a closed-ended control description is a higher bar to surmount. The issue here is not one of insufficient detail.[8] The issue is qualitative rather than quantitative: the control operations are not being described in the right *kind* of way.

My purpose here has not been to call into question the merit of the authors' work. All of the models that Braver et al.'s (2014) review discusses may be perfectly valid models that increase our understanding, allow for accurate predictions, and pave the way to future discoveries. But the models discussed here should not be counted as *mechanistic explanations*, in the New Mechanist sense, of their target phenomena. Instead, they may at best be counted as mechanism sketches (in the terminology of Craver 2006) since they identify the organization of and causal pathways between certain components as well as identify the operations of some of those components in mechanistic terms.[9] Further, due to the complexity of the anatomy and functions of the brain, we should not blame neuroscientists for the fact that they do not yet offer us complete mechanistic explanations of high-level brain functions like cognitive control: brain science is only starting to answer basic questions about how high-level functions are implemented in the brain like goal representations, the execution of plans, and thinking.

The Multivocality of "Mechanism"

But if mechanistic explanation of high-level brain functions is in such short supply, why do we find review articles with titles like "Mechanisms of Motivation–Cognition Interaction" advertising "specific mechanisms" and "mechanistically explicit" accounts? Instead of arguing that systems neuroscientists (or cognitive neuroscientists, and so on) are misusing the word "mechanism," I think we must recognize the presence of two quite distinct senses of the word. As Hommel (2020) and van Bree (2024) have pointed out, when systems neuroscientists say that they have identified a "mechanism" for how X "mediates," "modulates" (or some other verb) Y, what they often mean is little more than that some physical conditions necessary for X to be able to mediate, modulate (and so on) Y are present.[10] In other words, what goes under the name "mechanistic explanation" in systems neuroscience (particularly in the study of high-level cognitive, affective, and motivational phenomena) is perhaps a form of causal explanation, but often not a *mechanistic* explanation (or at best an incomplete mechanistic explanation) in the stricter sense. This is why, so long as a neuroscientist has found a medium by which a causal influence may be transmitted from one locus or system to another (for example, a neurotransmitter system or a neural circuit),

[8] As discussed in the section "The 'Black Box' Argument," abstraction from details is not, in itself, an obstacle for a model or explanation to count as mechanistic (Craver and Kaplan 2020).

[9] Though in the final section, I will suggest a reason why they might not count as mechanism sketches either.

[10] Krakauer et al. provide a list of such verbs commonly used by neuroscientists, which they call "filler verbs" (2017, p. 486).

that tends to be regarded as sufficient for having a "specific" mechanism or having made a mechanism "explicit."

This is very different from the New Mechanist's (and in particular, the molecular biologist's) usage of "mechanism" as an account that includes the parts, their operations, and how they are organized, sufficient to understand *how* the phenomenon comes about in terms of operations that are more basic—and leave less questions unanswered—than the function being explained.[11] These distinct senses of "mechanism" also imply the presence of distinct senses of "mechanistic explanation" with distinct norms.[12] While mechanistic explanation in the stricter New Mechanist sense prevents the inclusion of open-ended control operations, mechanistic explanations in the weaker systems neuroscience sense can include components described in terms of open-ended control operations. But if that is true, then why count such explanations as "mechanistic" at all? And if we should count such explanations as "mechanistic," then why not count all mere mechanism sketches, or even those purported mechanistic explanations that rely on filler terms to do explanatory work, as mechanistic explanations?

In a domain where mechanistic explanations in the New Mechanist sense are often not yet possible due to the early state of the science, investigators must seek out those explanations that will directly lay the groundwork for the eventual possibility of mechanistic explanations in the stronger sense, and they need a way of categorizing those explanations apart from explanations that are not directed towards mechanistic understanding in the same kind of way. Mechanistic understanding in the stronger sense requires identifying parts, identifying operations, identifying routes by which the parts can influence each other, and the organization and arrangement of them in space and time. In a massively complex structure like the brain with vast numbers of control systems that are only starting to be catalogued, finding where they are located and how they can influence each other is sometimes all that can be done, and explanations that shed light on such questions may be called "mechanistic" when the state of a given science or subfield is such that that is the closest we can come to a mechanistic explanation in the stronger sense (and so long as the provisional status of such open-ended characterizations of control are kept in mind, to recall Craver's (2006) warning about filler terms cited above). The alternative would be to constantly qualify explanations as "sketches" or as "incomplete," and so on, which would likely add more confusion.

We might therefore adopt a modified understanding of "mechanism" and "mechanistic" in the context of a scientific field that is at a state where mechanistic explanation in the New Mechanist sense is frequently not yet attainable, and should not be considered the standard for gauging scientific progress.[13] The lesson for philosophers, however, is that it is important to keep the terminology straight and understand that the goals and norms of such language are not the same across fields. "Mechanism" and "mechanistic explanation" are not univocal terms in science. In the long run, the goals and norms may coincide. Eventually, systems neuroscience may reach a point where satisfactory mechanistic explanations will require reference only to closed-ended control. But the point here is that "mechanistic explanation" can involve different goals and norms depending on the state of the field of study.

Conclusions

I have argued that we should distinguish between two ways control operations can be described: closed-ended control versus open-ended control. Closed-ended control behaves in a well-defined and predictable manner, characterized by a fully articulated input–output relation. The input and output conditions are specified as a determinate set of possibilities, and the systematic relationship that maps inputs to outputs is made fully explicit in non-normative terms, leaving no mystery about how the controller will respond in given circumstances. Conversely, open-ended control implies that the controller possesses the ability to exert influence across an open-ended range of circumstances, maintaining consistency of outcomes (by selecting whatever response is *appropriate* for achievement of those outcomes) but lacking an explicitly defined input–output mapping relation.[14]

A key difference is that when treating a component as capable of open-ended control, you are making a normative claim: you are saying that the component, itself, has a kind of reliability that goes beyond causal regularity and is something like a rudimentary or proto-version of instrumental rationality or practical wisdom (without necessarily

[11] Successful mechanistic explanations do often leave researchers with more questions than they started with (for example, an explanation E_1 may raise questions about the lower-level mechanisms that implement the operations cited by E_1, how the components managed to get there, and so on). But an explanation that does its job should leave less questions remaining about how its target phenomenon comes about at the compositional level, at the level of detail, and under the range of conditions, that the explanation is concerned with.

[12] Krakauer et al. similarly argue that "a more pluralistic conception of mechanistic understanding" is needed for neuroscience (2017, p. 488).

[13] van Bree (2024) argues to the contrary that cognitive neuroscience should *now* be held to the stricter New Mechanist standard of explanation.

[14] A control system might incorporate some degree of stochasticity into its input–output relation, rendering its outputs probabilistic rather than deterministic. This could still count as closed-ended in my terminology if the probabilities are treated as determined by non-normative factors.

implying the presence of cognition, a "homunculus," or the processing of representations). The component can be relied on (within some bounds) to select the appropriate means of getting the job done, *because* it is the appropriate means. The appropriateness of the means is then helping to do explanatory work, not the fact that antecedent operations and conditions in the mechanism and the specified input–output relation of the controller were causally sufficient for it.[15] This (sub-cognitive analogy to) instrumental rationality or practical wisdom is taken for granted or even treated as brute or primitive, rather than treated as something requiring explication in mechanistic terms. The inclusion of such a component in an explanation introduces an unreduced *teleological* element and, to that extent, prevents the explanation from qualifying as fully mechanistic (in the New Mechanist sense). Including a controller as a component in a mechanism, and describing its operation merely as that it maintains a setpoint, is not allowable as a form of abstraction or idealization, because the input–output relation (that is, *what* the controller does to maintain the setpoint and *when*) is not merely a detail to be included or excluded; it is what makes the operation a *mechanistic* (rather than teleological) one in the first place. The input–output relation itself can be described at varying levels of abstraction or detail, just so long as it is described as a determinate input–output relation and not as an open-ended capacity to exert whatever influence is needed for a certain condition to obtain. The upshot is that explanations that enlist control operations to do explanatory work can count as mechanistic in the strict sense only if such control operations are closed-ended, not open-ended.

In a scientific field such as systems neuroscience, it is not yet possible to provide mechanistic explanations for many phenomena in terms only of closed-ended rather than open-ended control. In such cases, scientists provide what they describe as "mechanistic explanations" that would not quite rise to the standard of what goes by "mechanistic explanation" according to New Mechanists. Arguably, these should not be characterized as what Machamer et al. (2000) refer to as "mechanism schemata," since a schema differs from a mechanistic explanation only in terms of the *amount* of detail it contains, not the *kind* of explanatory work (that is, mechanistic or not) it helps to do. Similarly with what Machamer et al. (2000) refer to as "mechanism sketches": again, these differ from mechanistic explanations in the New Mechanist sense by the amount of detail included or excluded. The difference between a sketch and a schema is the *reason* information is missing. In the case of a sketch, it is missing due to gaps in the knowledge of researchers, whereas in the case of a schema, details are intentionally missing so that the schema can play certain roles in scientific investigation (Machamer et al. 2000, pp. 16–18). By contrast, "mechanistic explanation" in the weak, non-New-Mechanist sense I have laid out may appeal to certain factors that prevent the explanation from being fully mechanistic, describing operations in terms of the consistent appropriateness of the outcome they (somehow) reliably ensure under open-ended circumstances, rather than a determinate mapping from circumstances to specific responses.

As a final note, I want to clarify that the distinction between closed-ended and open-ended control is intended as a distinction between how a controller is *characterized*, not as a metaphysical distinction about ways for a controller to *be* in itself. In other words, the distinction is not intended to be a *metaphysical* or "ontic" distinction. It is instead a semantic (or what some like to call an "epistemic") distinction. With sufficient knowledge (for example, the knowledge of Laplace's Demon), any controller that can be understood in terms of open-ended control might potentially be understandable in closed-ended terms, though the question of whether or not there exist control systems that cannot *in principle* be described in terms of closed-ended control is orthogonal to the points being made in this article.

Acknowledgments Thanks to William Bechtel, Leonardo Bich, Daniel S. Brooks, Dan Burnston, Fermín Fulda, Linus Huang, Robert McCauley, Andrew Richmond, and Denis Walsh for helpful comments and discussion.

Funding Funding for this project was provided by grant 61369 from the John Templeton Foundation.

Declarations

Competing Interests The author has no competing interests to declare that are relevant to the content of this article.

References

Bechtel W, Abrahamsen A (2005) Explanation: a mechanist alternative. Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 36:421–441. https://doi.org/10.1016/j.shpsc.2005.03.010

Bechtel W, Abrahamsen A (2011) Complex biological mechanisms: Cyclic, Oscillatory, and autonomous. In: Hooker CA (ed) Philosophy of complex systems. North Holland, Amsterdam, pp 257–285. https://doi.org/10.1016/B978-0-444-52076-0.50009-2

Bechtel W, Bich L (2021) Grounding cognition: heterarchical control mechanisms in biology. Philos Trans R Soc Lond B Biol Sci 376:20190751. https://doi.org/10.1098/rstb.2019.0751

Bechtel W, Richardson RC (1993) Discovering complexity: decomposition and localization as strategies in scientific research. Princeton University Press, Princeton

[15] Some might argue (for example, McGrath 2005, p. 140) that a regular input–output relation is, by itself, sufficient for a system to count as embodying a norm. Even if true, this additional fact could not be employed to do explanatory work in a purely mechanistic model.

Braver TS (2012) The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn Sci 16:106–113. https://doi.org/10.1016/j.tics.2011.12.010

Braver TS, Gray JR, Burgess GC (2007) Explaining the many varieties of working memory variation: dual mechanisms of cognitive control. In: Conway A, Jarrold C, Kane M (eds) Variation in working memory. Oxford University Press, New York, pp 76–106. https://doi.org/10.1093/acprof:oso/9780195168648.003.0004

Braver TS, Krug MK, Chiew KS, Kool W, Westbrook JA, Clement NJ, Cognitive et al (2014) Affect Behav Neurosci 14:443–472. https://doi.org/10.3758/s13415-014-0300-0

Craver CF (2006) When mechanistic models explain. Synthese 153:355–376. https://doi.org/10.1007/s11229-006-9097-x

Craver CF (2014) The ontic account of scientific explanation. In: Kaiser MI, Scholz OR, Plenge D, Hüttemann A (eds) Explanation in the special sciences: the case of biology and history. Springer, Dordrecht, pp 27–52. https://doi.org/10.1007/978-94-007-7563-3_2

Craver CF, Kaplan DM (2020) Are more details better? On the norms of completeness for mechanistic explanations. Br J Philos Sci 71:287–319. https://doi.org/10.1093/bjps/axy015

Cummins RC, Poirier P (2004) Representation and indication. In: Clapin H, Staines PJ, Slezak P (eds) Representation in mind: new approaches to mental representation. Elsevier, Amsterdam, pp 21–40. https://doi.org/10.1016/B978-008044394-2/50005-1

Dennett DC (1973) Mechanism and responsibility. In: Honderich T (ed) Essays on freedom of action. Routledge and Kegan Paul, London, pp 157–184

Dretske F (1981) Knowledge and the flow of information. Bradford, Cambridge

Glennan SS (1996) Mechanisms and the nature of causation. Erkenntnis 44:49–71. https://doi.org/10.1007/bf00172853

Glennan SS (2017) The new mechanical philosophy. Oxford University Press, Oxford. https://doi.org/10.1093/oso/9780198779711.001.0001

Hommel B (2020) Pseudo-mechanistic explanations in psychology and cognitive neuroscience. Top Cogn Sci 12:1294–1305. https://doi.org/10.1111/tops.12448

Kauffman SA (1971) Articulation of parts explanation in biology and the rational search for them. In: Buck RC, Cohen RS (eds) PSA 1970: in memory of Rudolf carnap. Reidel, Dordrecht, pp 257–272. https://doi.org/10.1007/978-94-010-3142-4_18

Keener J, Sneyd J (2009) Mathematical physiology II: systems physiology, 2nd edn. Springer, New York. https://doi.org/10.1007/978-0-387-79388-7

Khoo MCK (2018) Physiological control systems: analysis, simulation, and estimation, 2nd edn. Wiley, Hoboken. https://doi.org/10.1002/9781119058786

Krakauer JW, Ghazanfar AA, Gomez-Marin A et al (2017) Neuroscience needs behavior: correcting a reductionist bias. Neuron 93:480–490. https://doi.org/10.1016/j.neuron.2016.12.041

Machamer PK, Darden L, Craver CF (2000) Thinking about mechanisms. Philos Sci 67:1–25. https://doi.org/10.1086/392759

MacKay DM (1964) Cybernetics. In: Brierley J (ed) Science in its context: a symposium with special reference to Sixth-Form studies. Heinemann, London, pp 305–318

McGrath S (2005) Causation by omission: a dilemma. Philos Stud 123:125–148. https://doi.org/10.1007/s11098-004-5216-z

McKenna M (2017) Reasons-Responsive theories of freedom. In: Timpe K, Griffith M, Levy N (eds) The Routledge companion to free will. Routledge, New York, pp 27–40. https://doi.org/10.4324/9781315758206-9

McKinley MJ, Denton DA, Ryan PJ et al (2019) From sensory circumventricular organs to cerebral cortex: neural pathways controlling thirst and hunger. J Neuroendocrinol 31:e12689. https://doi.org/10.1111/jne.12689

Milsum JH (1966) Biological control systems analysis. McGraw-Hill, New York

Pattee HH (1971) Physical theories of biological co-ordination. Q Rev Biophys 4:255–276. https://doi.org/10.1017/s0033583500000640

Pattee HH (1972) The nature of hierarchical controls in living matter. In: Rosen R (ed) Foundations of mathematical biology, vol 1. Subcellular Systems. Academic, New York, p pp 1-22

Pattee HH (1973) The physical basis and origin of hierarchical control. In: Pattee HH (ed) Hierarchy theory: the challenge of complex systems. Braziller, New York, pp 71–108

Shenhav A, Botvinick MM, Cohen JD (2013) The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron 79:217–240. https://doi.org/10.1016/j.neuron.2013.07.007

Shinar G, Milo R, Martinez MR, Alon U (2007) Input–output robustness in simple bacterial signaling systems. Proc Natl Acad Sci U S A 104:19931–19935. https://doi.org/10.1073/pnas.0706792104

Shoemaker S (1980) Causality and properties. In: van Inwagen P (ed) Time and cause: essays presented to Richard Taylor. Reidel, Dordrecht, pp 109–135. https://doi.org/10.1007/978-94-017-3528-5_7

Sourjik V, Wingreen NS (2012) Responding to chemical gradients: bacterial chemotaxis. Curr Opin Cell Biol 24:262–268. https://doi.org/10.1016/j.ceb.2011.11.008

Sultan SE, Moczek AP, Walsh DM (2022) Bridging the explanatory gaps: what can we learn from a biological agency perspective? Bioessays 44:2100185. https://doi.org/10.1002/bies.202100185

van Bree S (2024) A critical perspective on neural mechanisms in cognitive neuroscience: towards unification. Perspect Psychol Sci 19:993–1010. https://doi.org/10.1177/17456916231191744

Walsh DM (2015) Organisms, agency, and evolution. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9781316402719

Willems JC (1995) Control as interconnection. In: Francis BA, Tannenbaum AR (eds) Feedback control, nonlinear systems, and complexity. Springer-, London, pp 261–275. https://doi.org/10.1007/BFb0027681

Winning J, Bechtel W (2018) Rethinking causality in biological and neural mechanisms: constraints and control. Minds Mach 28:287–310. https://doi.org/10.1007/s11023-018-9458-5