Erkenn (2009) 71:233–251
DOI 10.1007/s10670-008-9125-y

ORIGINAL ARTICLE

A Modified Conception of Mechanisms

Phillip J. Torres

Received: 13 March 2008 / Accepted: 18 September 2008 / Published online: 22 October 2008

Abstract In this paper, I critique two conceptions of mechanisms, namely those put forth by Stuart Glennan (Erkenntnis 44:49–71, 1996; Philosophy of Science 69:S342–S353, 2002) and Machamer et al. (Philosophy of Science 67:1–25, 2000). Glennan's conception, I argue, cannot account for mechanisms involving negative causation because of its interactionist posture. MDC's view encounters the same problem due to its reificatory conception of activities—this conception, I argue, entails an onerous commitment to ontological dualism. In the place of Glennan and MDC, I propose a "modified conception" of mechanisms, which (a) obviates the problem of negative causation by reinterpreting MDC's activities according to a "descriptivist" account, and (b) avoids MDC's problem by postulating a monistic ontology of entities. Thus, by solving these problems, my modified conception offers a cogent, more adequate alternative to Glennan's and MDC's conceptions of mechanisms.

1 The New Mechanical Philosophy

The last 15 years has seen the burgeoning of a new field in philosophy of biology focused on *mechanisms*. As many philosophers have pointed out, mechanisms occupy a central position in the mosaic of biological theory, and as such deserve the philosophical attention of those interested in understanding the nature of biology. A

P. J. Torres (✉)
Department of Philosophy, Harvard University, Emerson Hall 209a, 25 Quincy Street, Cambridge, MA 02138, USA
e-mail: philosophytorres@gmail.com; ptorres@fas.harvard.edu

survey of the literature on mechanisms reveals at least two impetuses behind the philosophical analysis of mechanisms:

(i) A better understanding of mechanisms—their componency, internal structure, hierarchical relations, etc.—means a better understanding of *mechanistic explanation*. The mechanistic model of explanation represents an appealing alternative to classical covering-law (CL) models, which construe explanations as logical arguments whereby a description of the phenomenon (the explanandum) is deduced from premises (the explanans) that include at least one law of nature. Laws of nature have in turn been characterized of as universal, exceptionless, counterfactual supporting generalizations with a complex logical form.[1]

Two observations lead to the conclusion that CL models are inadequate for the biological sciences: first, the biological sciences have a dearth of—if any at all—lawful generalizations; and second, biology is not a merely descriptive discipline, but provides explanations of the explananda phenomena within its various subdomains. These explanations tend to be *mechanistic* in character rather than *nomological*, and proceed by describing underlying mechanisms responsible for the phenomena of interest. For example, molecular biologists explain protein synthesis by describing the "DNA → (transcription) RNA → (translation) Protein" mechanism in cells; geneticists explain the replication of DNA by describing the DNA polymerase mechanism responsible; neurobiologists explain long-term potentiation (LTP) by describing the NMDA channel mechanism of LTP; and so on.

What emerges from these examples is an identity relation between mechanism descriptions and mechanistic explanation. In other words, the aphorism might go: *mechanism descriptions explain phenomena, and mechanistic explanations describe mechanisms*. Quite naturally, this leads to the question: What *are* mechanisms? What distinguishes a good mechanism description—and therefore a good mechanistic explanation—from a bad one?

(ii) A second reason for growing philosophical interest in mechanisms pertains to causation. Since David Hume propounded his causal skepticism, whereby causation was conceived as a mere "connecting principle" (Hume 2006, p. 23), i.e., an associative relation enabling the mind to transcend "the evidence of our memory and senses" (quoted from Morris 2008), philosophers have struggled to give an adequate metaphysics of causation. Observing that the relationship between mechanisms and their phenomena is intrinsically causal, some philosophers have attempted to formulate reductive accounts of causation in purely mechanistic terms. Stuart Glennan, for example, proposes a "mechanical theory of causation" that he argues offers a partial solution to Hume's connection-conjunction (CC) problem by weakening it to "not a universal one" (Glennan 1996, p. 68). (In Glennan's terminology, "connections" are genuine causes—i.e., *X* causes *Y*—whereas "conjunctions" are not—i.e., *X* and then *Y*.)

Glennan predicates his mechanical theory on a metaphysical bifurcation of reality into (a) the level of fundamental physics, and (b) the non-fundamental level of all else. While Glennan concedes that "Hume's [CC problem] still remains" on

[1] See Pietroski and Rey (1995).

the former level, since "we can observe certain regularities, but we cannot offer an explanation of why those regularities obtain" (Glennan 1996, p. 68), he theorizes that *X* causes *Y* on the latter level when and only when a mechanism links the two variables (which stand for property-change events).

In other words, it is by virtue of a mechanism that causal connections are, on all levels of nature other than that of fundamental physics, metaphysically distinct from conjunctions. In *Explaining the Brain* (2007a), Carl Craver argues that solving the CC problem by facilely relegating it to the level of fundamental physics is unconvincing. Making matters worse, Craver states, a tension emerges in Glennan's mechanical theory between its anti-fundamentalist stance about causation (i.e., that causes exist—in the form of mechanisms—on non-fundamental levels; see Craver 2007a, pp. 86–93) and its reductionistic posture (which I discuss more in Sect. 2). These problems are sufficient, Craver claims, to render Glennan's mechanical theory inadequate as a solution to Hume's CC problem.

Putting aside the apparent flaws of Glennan's mechanical theory, to give a reductive account of causation in terms of mechanisms first requires an answer to the question: What *are* mechanisms? By virtue of what are mechanisms causal agents of change? The formulation of both (i) a model of mechanistic explanation and (ii) an account of causation therefore presupposes an adequate *conception of mechanisms*, i.e., an account of what is meant by "mechanism." The "new mechanical philosophy," as Skipper and Millstein (2005) have termed it, consists of attempts to devise an adequate conception of mechanisms that would prescribe norms for the conceptualization of mechanisms. Such attempts include those by Stuart Glennan (1996, 2002), Machamer et al. ('MDC' hereafter) (2000), and William Bechtel (2006).[2]

Although these conceptions of mechanisms differ importantly in their details, all fall within the *systems tradition*, which holds that one "construes explanation as a matter of decomposing systems into their parts and showing how those parts are organized together in such a way as to exhibit the explanandum phenomenon" (Craver 2007a, p. 109). The act of "showing" here refers to the construction of a mechanism description, which one completes in accordance with the normative conception that one espouses. In sum, the adequacy of a mechanistic explanation is dependent upon the adequacy of the mechanism description with which it is identical, and this description is in turn dependent upon the adequacy of the conception of mechanisms that normatively guided its construction.

In this paper, I will argue that neither Glennan's nor MDC's accounts of mechanisms are adequate, but that a synthesis combining the best features of each yields an adequate *modified conception*. Let us begin with Glennan's interactionist conception of mechanisms.

[2] I do not discuss William Bechtel's parts/operations conception in this paper, although it is, at least in certain respects, very similar to the modified conception that I enunciate in Sect. 5.

2 Glennan's Complex-Systems Approach and Woodward's Interventionism

Glennan (2002) defines a mechanism as:

> A complex system that produces [a] behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations. (Glennan 2002, p. S344)

This definition differs slightly from Glennan's 1996 definition in that it replaces the characterization of intra-mechanism interactions as "direct, causal laws" (Glennan 1996, p. 52) with "invariant, change-relating generalizations." Glennan borrows these terms and their corresponding concepts from Jim Woodward's manipulability theory of causation, which I refer to as *interventionism*.

In contrast to Glennan's 1996 *reductive* theory of causation (discussed in Sect. 1), Woodward's interventionism is a *non*-reductive account of causation, characterizing causes in causal (rather than non-causal) terms. Since Glennan's 2002 mechanism conception employs Woodward's manipulability theory, it follows that to understand Glennan's account one must understand Woodward's theory. A familiarity with interventionism is also necessary for an understanding of my modified conception, which similarly incorporates Woodward's apparatus into its theoretical framework. In the following paragraphs, I summarize the relevant features of Woodward's interventionism.

To begin, recall that the biological sciences are devoid of laws of nature, conceived of as generalizations holding without exception, under all conditions, and supporting counterfactual conditionals (cf. *ceteris paribus* laws,[3] accidental generalizations[4]). More generally, Woodward points out that the special sciences (viz., all domains of scientific inquiry minus physics) have relatively few such generalizations, yet nevertheless seem to provide robust explanations of the phenomena within their respective domains. Based on these observations, Woodward proposes the following criterion for a generalization to count as explanatory: it must be *invariant*.

Thus, on Woodward's account, rather than talk of "laws of nature," philosophers ought to speak of "invariant generalizations." Woodward spells out the notion of invariance in terms of *interventions*, or exogenous manipulations of component parts in a causal system (e.g., a mechanism), effectuated either by human agency (a scientist) or some natural occurrence (a neoplasm). These manipulations serve to change the value of a variable, where variables are the relata of causal relations.[5] In sum, invariance is a property of generalizations that hold stable over some "domain" of interventions.

Like traditional conceptions of laws, which are said to support counterfactual conditionals, Woodward's interventions also support counterfactuals, although of a

[3] See Pietroski and Rey (1995) for discussion.

[4] Incidentally, Woodward argues that the dichotomy between laws and accidental generalizations is spurious and ought to be abandoned.

[5] As Craver points out, the use of variables as causal relata is advantageous because a wide variety of possible relata—events, processes, objects, etc.—can be converted "without loss" into talk of variables (Craver 2007a, p. 95).

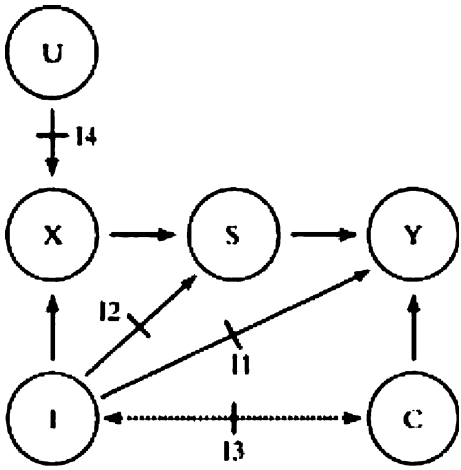

Fig. 1 Craver's diagram of "an ideal intervention on *X* with respect to *Y*" (Craver 2007a, p. 97). 'U' stands for other causes, and 'S' stands for a causal intermediate. The arrows perpendicularly bisected by the short lines signify possible causal relations that should not obtain. These are the conditions that the formula assumes hold. See Craver (2007a, pp. 93–98)

peculiar sort. Specifically, interventions satisfy the antecedents of so-called "active counterfactuals"; Woodward's interventionism thus explicates causation in terms of *active counterfactual dependence*. In other words, as an abstract formula: c (rather than $\sim c$) causes e (rather than $\sim e$) when an ideal intervention could be made, given some set of conditions C_1, C_2, $C_3 \ldots C_n$ (see Fig. 1), to change the value of c such that the value of e would also change.[6] In this way, e exhibits active counterfactual dependence with respect to c.

One might object that in some counterfactuals, satisfying the antecedent is not physically or technologically possible. For example, astronomical interventions to determine causal relationships between planetary orbits and the sun's gravitational field pose serious practical problems for the interventionist, since moving the sun's position in space relative to the planets in the solar system is impossible. Woodward surmounts this problem by stipulating that interventions need only be *ideal*. Thus, given the counterfactual: "If property X of entity E_1 were changed from x_1 to x_2, then property Y of entity E_2 would change from y_1 to $y_2 = f(x_2)$" (adapted from Craver 2007a, p. 202), one must minimally show that when an *ideal intervention* is made on E_1, it brings about the specified change in E_2.

Returning to Glennan's conception of mechanisms, we should now have a better idea of what Glennan means by characterizing intra-mechanism interactions in terms of invariant, change-relating generalizations. (The change-relating part, which I did not discuss above, simply means that it is possible to wiggle one entity, E_2, by wiggling the other, E_1.) But we have not yet answered the question: What *are* interactions? To begin, note an ambiguity in this question between a *metaphysical* and *definitional* interpretation: one could be asking for an account of the ontological status of interactions, or one could be requesting a set of necessary and sufficient conditions for some causal event X to count as an interaction, respectively.

[6] I follow Craver in understanding causal relata as contrasts, and thus the adicity of causal relations to be four (Craver 2007a, pp. 82–83).

Regarding the metaphysical interpretation, Glennan is not explicit about the ontology of his conception, but all indications point to a monistic metaphysics. For example, Glennan explicitly rejects MDC's postulation of activities as constituting a distinct ontological category[7] (see Sect. 3); furthermore, MDC characterize Glennan as a substantivalist (MDC 2000, p. 5). Thus, we are justified in assuming that Glennan's ontology posits entities as ontologically basic with interactions serving a solely descriptive purpose in mechanism models (by specifying which property changes of what parts cause which property changes of what other parts; see below). A "mechanism model" on Glennan's (2002) account is just a description of both the mechanism and its behavior.

Let us move on to the definitional interpretation of the question: What are interactions? Glennan stipulates in both his 1996 and 2002 papers that an interaction signifies "an occasion on which a change in a property of one part brings about a change in a property of another part" (Glennan 2002, p. S344). As James Tabery points out, though, the collocation 'brings about' is causal, and rather vague at that. Tabery therefore objects that mechanism descriptions constructed within the interactionist framework will yield descriptive (and thus explanatory) "black boxes" (Tabery 2004, p. 10), since simply specifying which property changes bring about which other property changes leaves out *how* the bringing about occurs.

The use of 'how' here is ambiguous, though. It could refer to a *description* of the causal relation between two entities, e.g., binding rather than pushing or docking, or it could refer to the *physical agent* that actually does the bringing about, e.g., a protein, a person, or a planet. These are two epistemically distinct pieces to Tabery's puzzle: one can know *what*, e.g., a protein of kind *A*, brings about the observed changes without knowing exactly *how* protein A does this, e.g., by binding rather than pushing or docking. And conversely, one can know *how*, e.g., by binding, the observed changes are brought about without knowing exactly *what* brings them about, e.g., a protein of kind *A* rather than *B*. Thus, Tabery's objection has two aspects to it, one descriptive and the other ontic.

With respect to the ontic aspect, Glennan's conception offers an explicit answer: mechanisms consist of interacting parts mediated by lower-level mechanisms, which themselves consists of interacting parts mediated by even lower-level mechanisms, and so on until one bottoms out at the level of fundamental physics. This is the reductionistic feature of Glennan's account mentioned above, which Craver sees as theoretically incompatible with Glennan's anti-fundamentalism.[8] Suffice it to say that Glennan's conception answers one aspect of Tabery's black box objection (the cogency of this answer is another matter).

So what about the descriptive aspect? Although Glennan does not explicitly require that descriptions of intra-mechanism interactions include specifications of the causal processes that bring about the property changes of interest, complete mechanism descriptions would no doubt include such information. The reason is

[7] Personal communication.

[8] As I suggest below, Glennan (2002) appears to renounce his 1996 anti-fundamentalism about causation by adopting Woodward's manipulability theory of causation, although the two are not necessarily incompatible.

that Glennan's conception of mechanisms, and indeed all mechanism conceptions, occupy a level of abstraction far above the concrete descriptions that scientists put forth in textbooks and journals.

As the description becomes more concrete, vague causal terms are replaced by less vague terms, and so on until the ideally complete description includes precise language. As Craver points out, scientists often use terms like 'inhibit',[9] 'encode', 'cause', 'produce', 'process', and so on as "filler terms" until further research can provide more information, thereby filling in the descriptive-explanatory lacunas (Craver 2007a, p. 113). Thus, the black box problem (in which a mechanism description fails to say *how* the property changes of one entity bring about the property changes of another entity) is not a problem for Glennan's conception of mechanisms to solve; rather, it is one for scientists to solve in the laboratory.

Glennan's complex-systems conception thus has an answer for both the descriptive and metaphysical aspects of Tabery's bipartite objection. In sum, the descriptive aspect entails specifying what sort of causal event brought about the observed property changes (*binding* verses *pushing* interactions), and is the task of scientists to discover through empirical experimentation. The metaphysical aspect, on the other hand, entails specifying the physical substrate underlying those changes, which Glennan's conception postulates as a lower-level mechanism (except on the level of fundamental physics). In Sect. 4, I argue that the interactionist posture of Glennan's conception engenders an unforeseen problem, one that brings the entire theory down. Then, in Sect. 5, I put forth my modified conception, which selectively extracts the good features of Glennan's account, while discarding the problematic parts, thereby avoiding the pitfalls of exclusivist interactionism.

3 MDC's Dualistic Conception

In contrast to Glennan's monistic conception of mechanisms, MDC characterize mechanisms as:

> Entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. (MDC 2000, p. 3)

Many objections have been raised against MDC's characterization. For example, Skipper and Millstein (2005) discover a number of problems in applying it to the Darwinian mechanism of natural selection. The selection mechanism, Skipper and Millstein argue, (a) is irregular rather than regular, (b) fails to display the organization that MDC require, and (c) since sometimes its "relevant activities" cannot be specified, it fails to exhibit *productive continuity* between stages (Skipper and Millstein 2005, p. 335). Such continuity, according to MDC, is essential for the

[9] As I discuss below, following Craver I consider instances of inhibition to be negative causes. Indeed, inhibition, in contrast to "omission," falls within the category of "prevention." Prevention occurs when *C* causes not-*E*; not-*E* in turn is identical to not-*C* that causes *E*. In this way, inhibition is causally negative. Neural networks exhibit inhibition when one neuron hyperpolarizes another, thereby *preventing* it from firing. See Craver (2007a, pp. 6, 80–81) for more.

intelligibility of a mechanism, and therefore for the mechanistic explanation of its phenomenon. (MDC seem to champion a psychological model of explanation, as Craver (2007b) affirms.) Other possible counterexamples can be adduced, such as the mechanism behind the tricarboxylic acid cycle, which is neither irregular nor regular, but *cyclical*.[10]

In MDC's defense, they explicitly circumscribe the domains in which their conception applies to those of molecular biology and neurobiology. (Although MDC's conception is normative, it is based on metascientific observations of how scientists in these two domains actually go about mechanistically explaining explananda phenomena; as such, it is descriptive as well.[11]) This gestures at the quandary in which mechanism conceptions find themselves: on the one hand, they ought to be general, insofar as generality is a virtue of hypotheses (Quine and Ullian 1978); on the other hand, the conception ought not be *so* general that its normative force is lost through abstraction.

For example, Paul Thagard (2006) proposes the following alternative to MDC's conception: "a mechanism consists of a group of parts that have properties and relationship[s] with each other that produce regular changes in those properties and relationships, as well as to the properties and relationships of the whole group" (Thagard 2006, p. 56). This account, which includes vague terms like "group of parts" and "relationships," seems so general and all-encompassing that its force as a normative conception of mechanisms is vitiated. Thus, it seems judicious to restrict the applicability of one's conception to particular domains of scientific inquiry, thereby achieving (or attempting to achieve) a balance between the competing desiderata mentioned above. Following MDC, this is precisely what my modified conception does.

My objections to MDC's conception are less *a posteriori* than Skipper and Millstein's; they pertain not to its "fit" with a known biological mechanism, but to its heavy ontological commitments. To put it bluntly, I am skeptical of MDC's activities, which (I will argue) are nothing other than *reified causes* dressed up in MDC's idiosyncratic phraseology. My argument for this interpretation consists of inferences from specific passages in MDC's 2000 paper. It presupposes, though, an understanding of the motivations behind MDC's ontological dualization. Thus, I begin with an examination of why MDC adopt a dualistic metaphysics, rather than the monistic position of Glennan's account.

MDC's ontology consists of entities and activities, and represents a synthesis of two monistic ontologies, namely those of (1) substantivalism, or substance metaphysics, and (2) processism, or process metaphysics. The first motivation for combining these two traditions is to neutralize the persistent philosophical difficulties associated with each, viz., the problem of accounting for entities in strictly processual terms [as "patterns of stability in a sea of process" (Rescher

[10] Lindley Darden (personal communication) argues that MDC's conception can account for cyclical mechanisms. The only difference between these and linear mechanisms is that one must more or less *arbitrarily* select start and termination conditions, e.g., the Krebs cycle starts with isocitrate rather than α-ketoglutarate, or succinate rather than fumarate.

[11] Craver similarly characterizes his project in *Explaining the Brain* as consisting of a descriptive, and then normative step (Craver 2007a, p. viii).

2002)], and the problem of accounting for processes in terms of entities[12] (MDC 2000, p. 5; Rescher 1996, 2002). The resultant *dualistic* ontology thus captures the "healthy philosophical intuitions" of substantivalism and processism, and puts forth a radical view "of the world as being *onticly composed* of entities and activities" (Machamer 2002; my emphasis).[13]

The second motivation behind MDC's dualism pertains to the criterion of descriptive adequacy. As MDC argue, it is a matter of fact that biologists construct descriptions of mechanisms that include specifications of entities *and* activities. For example, biologists emphasize in descriptions of the protein synthesis mechanism both the entities involved, such as DNA, RNA polymerase, mRNA, tRNA, rRNA, ribosomes, etc. as well as the activities involved, such as transcribing, translating, etc. (These activity verbs are, of course, rather vague as written here. As Sect. 2 argues, actual descriptions of protein synthesis often replace them with more specific, descriptively contentful terms, such as: "RNA polymerase *binds* to the DNA strand to *transcribe* it.")

Unlike Glennan's interactions, though, activity verbs (gerunds, etc.) have ontic referents in addition to their descriptive content. In other words, MDC identify the *how* with the *what*—that is, *they are one and the same*. This introduces the third motivation behind MDC's conception: to give an account of causation by postulating activities as *reified causes* (although in different terms). As Craver himself states, MDC's activities are "primitive" and "unanalyzable" components of mechanisms (Craver 2007b); I therefore I classify MDC's view as a version of *causal primitivism*. For the skeptical readers who understand MDC as propounding a more ontologically modest position, I provide the following argument:[14]

Premise 1: Recall that MDC's ontology represents a synthesis of substantivalism and processism, where entities correspond to substantial things and activities to processes (MDC 2000, pp. 4–8).[15] This being the case, MDC write that processists "reify activities" (MDC 2000, p. 5), which thereby suggests that MDC also reify activities. [I surmise that this is what Machamer (2002) means by activities being "ontic."] Consistent with the special reificatory status of activities, MDC talk of picking out activities according to their location in space and time, writing: "Traditionally one identified and individuates entities in terms of their properties and spatiotemporal location. Activities, likewise, may be identified and individuated by their spatiotemporal location" (MDC 2000, p. 5).

[12] Lindley Darden envisages a "completely static" and "frozen world" without activities (Darden 2006, p. 278; see footnote).

[13] Machamer states in the very next sentence: "This is the dualistic position that [MDC] put forth" (Machamer 2002).

[14] There is considerable confusion among philosophers about the ontological status of MDC's activities. Many seem to think, like Tabery, that activities do not constitute a distinct ontological category. This is simply false, as both careful readings of MDC's paper and personal communication with its authors verify. Indeed, as Darden puts it in personal communication, MDC's dualism is a *radical break* from the inveterate tradition of substantivalism.

[15] Although activities are slightly different in character than processes, the former is derivative of the latter.

Further evidence comes from MDC's definition of activities as "the producers of change" (MDC 2000, p. 3), a patently metaphysical claim. Indeed, activities are not merely descriptive of the kind of change that occurs, *but are actually responsible, in the causal sense, for the change itself*. MDC's view is therefore radically anti-reductionistic. An anomalous example will help to make this point clear: Take the activity of *manufacturing*. On MDC's view, 'manufacturing' ought not be reduced to lower-level constituent activities that describe the movement of workers, machines, and so on. Rather, the term refers to a "primitive" and "unanalyzable" [to quote Craver (2007b) again] ontic "thing" that actually *brings about* the production of goods. While MDC's position seems less implausible when applied to biological activities like binding, pushing, and docking, this example makes salient just how radical its ontological commitment to activities is.

Premise 2: MDC supplant talk of "causation" with talk of "productivity."[16] This terminological switch is the idiosyncratic feature of MDC's phraseology that I allude to above. Thus, to say of some activity that it is productive is simply to say that it is causal. From these premises we infer the following definition, based on MDC's original characterization of activities as "producers of change": *Activities are reified causes of change*. To recapitulate, this follows directly from (a) the interchangeability of 'causes' and 'producers', given their synonymy (premise 2), and (b) a conception of the definiendum as reified, given its nature as the ontological "offspring" of processes (premise 1).

As I discuss in Sect. 4, the postulation of intra-mechanism activities as reified causes involves a "break from parsimony" (to quote MDC 2000, p. 4) that problematizes MDC's conception, weighing it down with onerous ontological commitments.[17] These commitments, I argue, render it inadequate as an alternative to Glennan's complex-systems approach. In the next two sections, I turn to Glennan's interactionist conception, which I argue is boxological, i.e., it engenders a black box, at least when it describes thermodynamic mechanisms. As such, Glennan's conception is also inadequate. I then propose in Sect. 5 my modified conception, which I claim stands as an adequate normative conception of mechanisms.

4 Thermodynamic Phenomena and Negative Causation

I begin this section with a quick look at the semantic relation between Glennan's term 'interaction' and MDC's term 'activity'. In the sections above, I use *binding* as an example of both an interaction *and* an activity. This suggests that the extensions of these terms overlap. But are they *co*extensive, or is one term a subset of the

[16] This has been confirmed through personal communication. The idea, I think, was to disencumber the *notion* of causation from the heavy connotations (and whatever other semantic baggage) that the *term* carries with it, and to do this by introducing into the philosophical lexicon a novel term.

[17] MDC would agree, I think, that their dualist ontology is "onerous." Indeed, this is precisely why they feel the need, as they say, to "*justify* this break with parsimony, this dualism of entities and activities" (MDC 2000, p. 4; emphasis added), by arguing for their conception's ontic, descriptive, and epistemic adequacy.

other? In his 2004 paper, Tabery states that there are "no productive activities that are not property-changing interactions" (Tabery 2004, p. 9). In other words, by obversion: All activities are interactions. Tabery proceeds to argue that although some activities appear prima facie to be *un*-interactive, closer examination will reveal them to involve interactions after all.

To corroborate his claim, Tabery adduces the cyclohexane molecule, which undergoes a conformation change from the "chair" to the "boat" form. At first glance, the activity of switching forms appears un-interactive. But if one conceptualizes the molecule itself as a mechanism, Tabery contends, one can explain the conformation change phenomenon in terms of micro-interactions on the atomic level. Thus, by identifying the putative activity with a phenomenon produced by a lower-level mechanism, Tabery ostensibly shows that activities which at first seem un-interactive turn out to be interactive.

I do not here controvert Tabery's claims that the cyclohexane molecule (a) is not un-interactive, and (b) may nonetheless appear un-interactive. All I wish to say is that Tabery's example is tendentious: only a hasty generalization can take us from a *single* confirming instance to a universal generalization about *all* activities. Indeed, what one needs to do is scour the landscape of known mechanisms in search of an anomalous case, i.e., a mechanism that involves an *un*-interactive activity. I argue that such activities do in fact exist, and that one can find them in any mechanism that involves thermodynamic phenomena. Specifically, they are the activities that bridge enabling conditions and spontaneous processes.

Take the hackneyed example of long-term potentiation (LTP) as our paradigm (MDC 2000; Craver 2007a) (Fig. 2). The mechanism works as follows: under specific conditions, such as when a nearby synapse becomes strongly activated or when an action potential (AP) back-propagates from the soma, the postsynaptic membrane depolarizes. This causes NMDA channels (see Fig. 3) to release Mg^{2+} ions from their pores, which prevent ions from passing through the channels when

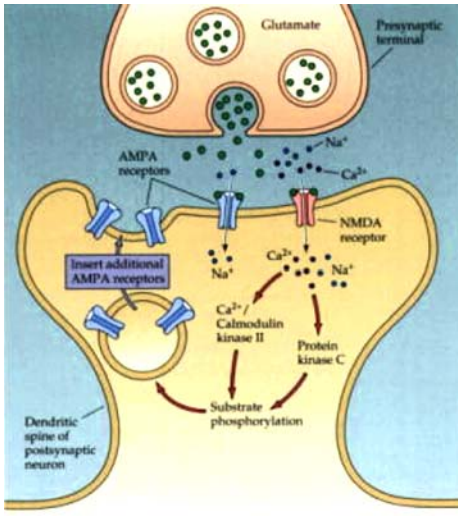

Fig. 2 The mechanisms of (early phase) long-term potentiation (LTP). (From Purves et al. 2004, p. 589)

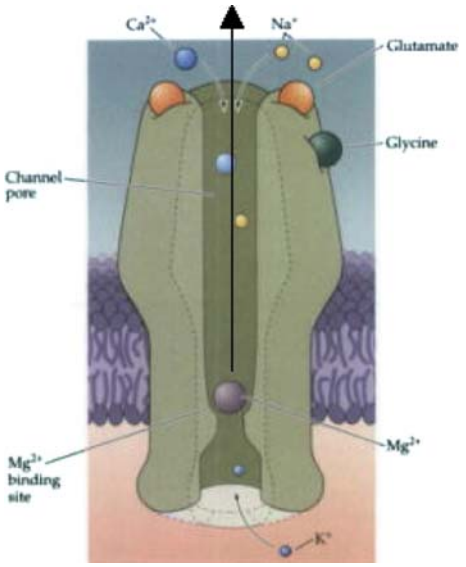

Fig. 3 The NMDA channel. (From Purves et al. 2004, p. 142)

intact. If glycine and glutamate are present, they bind to the NMDA channel, at which point Ca^{2+} diffuses into the dendrite. Ca^{2+} then acts as a second messenger, activating Ca^{2+}/calmodulin-dependent protein kinase type II (CaMKII) and protein kinase C (PKC). These then phosphorylate substrates that incorporate additional AMPA receptors into the postsynaptic membrane. The strength of the synapse consequently increases.

Most of this mechanism poses no problem for Glennan's interactionism. Indeed, neuroscientists often characterize steps in the LTP mechanism as interactions, such as when glycine and glutamate bind to the NMDA channel. But there are other steps, for example when the *removal* of Mg^{2+} from the NMDA channel pore brings about the *diffusion* of Ca^{2+}, that no neuroscientist would characterize as an interaction. Indeed, it is unimpeachable that the absence of Mg^{2+} ions in no way *interacts* with Ca^{2+} ions to effectuate their fluxion into the cell.

This is because diffusion is a spontaneous process (defined as a decrease in the amount of Gibbs free energy in the system) requiring only that certain enabling conditions obtain. In the case of LTP, the removal of the Mg^{2+} block *enables* Ca^{2+} ions to diffuse, thereby bringing about the change through an *absence* (rather than a presence). I want to make a connection here between instances of so-called "negative causation" and activities that have "thermodynamics as [their] source" (MDC 2000, pp. 13–14). MDC recognize such activities in their typology of activities as "energetic" (MDC 2000, pp. 13–14); energetic activities include those of diffusion, osmosis, effusion, convection, conduction, and so on.

Many mechanisms involving negative causes, e.g., absences, interferences, failures, non-occurrences, etc. also involve spontaneous processes.[18] Such

[18] Spontaneous processes may result not just from negative causes (such as in LTP), but also from positive causes; for example, the *presence* of salt may be responsible for the entropic movement of water across a cell membrane.

mechanisms pose a problem for Glennan's interactionist conception as well as MDC's notion of activities as immanent "producers of change," and this is because negative causes are apparently *un*-interactive[19] and certainly not immanent. As I explain below, my modified conception obviates these problems by adopting a "descriptivist" interpretation of activities according to which activity verbs (gerunds, etc.) are explanatory by virtue of their descriptive content.

To be clear, I have argued *contra* Tabery that while all interactions are instances of activities, not all activities are instances of interactions.[20] More specifically, activities that connect an absence (or interference, etc.) with a spontaneous process are not interactions but *un-interactive activities*. Furthermore, my claim should not be taken to imply that diffusion *itself* is un-interactive. (Indeed, the kinetic-molecular theory of gases states that particles in a gas are constantly colliding.) My claim is simply that in the LTP mechanism, for example, one can trace a *continuous* concatenation of interactions from the depolarization of the dendrite to the release of Mg^{2+} ions, as well as from the diffusion of Ca^{2+} to the incorporation of AMPA receptors into the postsynaptic membrane. But one *cannot* find an interaction linking these two events—and for no other reason than because *there is no interaction there to find*.

Let us move on to my descriptivist interpretation of activities.[21] Recall once again that MDC postulate activities as reified causes. They are, along with entities, "ontic" constituents of the world. I want to retreat from MDC's radical break with parsimony and adopt a characterization of activities similar to Glennan's notion of interactions. (Except that activities on my view encompass both *interactive activities*, i.e., interactions, and *un-interactive activities*, i.e., negative causes.)

Thus, activity verbs on the descriptivist interpretation have no ontic referent, but as I mention above are explanatory by virtue of their descriptive content—just like Glennan's interactions. In other words, activity verbs explain by specifying *how* a property change is brought about, without necessarily saying *what* does the bringing about. Indeed, as Glennan (1996) points out, descending to a lower-level (which is what specifying the *what* would entail) is unnecessary in most cases of explanation, and sometimes even explanatorily detrimental. All scientists care about when explaining an explanandum phenomenon is that they know, on a given mechanism level,[22] *how* the mechanism's parts are causally connected, e.g., by binding, pushing, or docking. It is the task of future scientific research to identify *what* lower-level (or possibly higher-level; see Darden 2006) mechanisms are causally

[19] To be clear, even in the case described in footnote 15, there still would be no interaction between the salt and the water.

[20] Darden says essentially the same thing: "Interaction between two entities is a *kind* of activity, conceived in a more entity-centered way and missing the productive nature of activities in general" (Darden 2006, p. 277; my emphasis). Although I disagree with the second statement that interactions miss the productive nature of activities (see Sects. 2 and 5), Darden and I concur that interactions are a subset of activities. Furthermore, an etymological parsing of 'interaction' corroborates my contention: 'inter-' meaning *between* and 'action' or *activity*.

[21] This interpretation is similar in many respects to Craver's "deflationary" view of activities (Craver 2007b).

[22] See Craver (2007a) for discussion about mechanism levels.

responsible for bringing about the observed property-changes, and so on until scientists can delineate a complete hierarchy of mechanisms.

Before proceeding, let me briefly address how the descriptivist interpretation accounts for instances of negative causation. As I state above, the trouble with Glennan's interactionism is that one cannot always identify a *continuous* concatenation of interactions from one intra-mechanism stage to the next. Consequently, mechanism descriptions constructed within the interactionist framework are black box engendering (although the black boxes here are not those that Tabery had in mind). Or, in MDC's phraseology, Glennan's account fails to show the productive continuity between stages in all mechanisms involving thermodynamic phenomena.

But MDC's conception also encounters difficulties, since talk of "negative activities,"[23] which Machamer once conceded might exist (Machamer 2002), is at least bizarre and at most incoherent. The situation is this: a *complete* description of the LTP mechanism, as found in standard neuroscience textbooks, includes the activities of Mg^{2+} removal and Ca^{2+} diffusion, as well as the activity of *allowing*. This activity is no less productive (i.e., change producing) than the other "positive" activities; indeed, it is the *allowance* of Ca^{2+} ions to diffuse into the neuron *via* the removal of Mg^{2+} ions that makes the mechanism work. An adequate conception of mechanisms, therefore, ought to account for this very important activity, not just the activities before and after it.

On MDC's view, considering the activity of allowing to be a reified cause is problematic because absences have no causal powers. On the other hand, not considering allowing to be a genuine activity (in MDC's sense of the term), while considering removal and diffusion to be genuine activities, would seem arbitrary and capricious. Since neither of these options is cogent, MDC's conception fares no better than Glennan's with respect to negative causation.

My descriptivist interpretation neutralizes these problems (a) by allowing one, first and foremost, to talk about negative causes (i.e., un-interactive activities), as well as (b) by rejecting the reificatory status that MDC confer to activities, thereby obviating the tension between negative causes as *absences* (or interferences, etc.) and MDC's activities as *presences*.

But how does the descriptivist interpretation define activities? We can derive a simple definition (albeit tentative) of activities from Glennan's definition of interactions. First, recall that Glennan defines an interaction as "an occasion on which a change in a property of one part brings about a change in the property of another part." We can thus define an activity as: *an occasion on which one or more entities brings about one or more property changes*. This definition includes cases in which (i) the property changes of one entity bring about those of another (interactions), and (ii) the property changes of a single entity serve as the enabling conditions for the spontaneous property changes of another entity (negative causation).

In the next section, I explore how a conception that incorporates this definition of activities can account for mechanisms involving thermodynamic phenomena

[23] That is, understood as reified causes.

without engendering the problems that stultify Glennan's and MDC's conceptions. It goes without saying, of course, that an adequate conception of mechanisms ought to permit one to construct non-boxological descriptions (and therefore good mechanistic explanations) of mechanisms that include spontaneous processes, especially since a large number of biological mechanisms involve diffusion, osmosis, etc.

5 A Modified Conception of Mechanisms

I propose the following modified conception of mechanisms:

> (MC) Mechanisms are complex systems composed of entities organized in space and time such that (i) through engaging in activities they produce a phenomenon, and (ii) the activities in which the mechanism's entities engage are characterizable in interventionist terms of direct, invariant, change-relating generalizations.

This conception has two distinguishing features worth making explicit: first, it postulates a monistic ontology of entities, and second, it accounts for instances of negative causation, which (I hope to have shown in Sect. 4) neither Glennan's nor MDC's conceptions do. But my conception also has many features in common with Glennan's and MDC's accounts. For example, it borrows from MDC talk of "activities" (although understood according to the descriptivist interpretation) and follows Glennan in employing Woodward's interventionism to characterize activities. As such, my modified conception represents a *synthesis* of Glennan and MDC, selectively extracting from each their theoretical strong points while discarding the problematic features.

As a synthesis, my modified conception ought to be distinguished from Tabery's (2004) "interactivity thesis," which proposes a dissimilar "synthesis" of Glennan and MDC. To begin, Tabery's proposes the neologism 'interactivity'[24]—a portmanteau of 'interaction' and 'activity'—which signifies "an occasion on which a change in a property of one part dynamically produces a change in a property of another part" (Tabery 2004, p. 12). On Tabery's view, philosophers ought to abandon talk of interactions and activities and instead speak of *interactivities*.

By adopting the language of interactivities, Tabery argues that mechanism descriptions can capture the "dynamicity" of mechanisms as producers of phenomena while (a) avoiding the black boxes that Glennan's interactions engender (I argue in Sect. 3 that Tabery's black box argument is ineffectual and misguided), and (b) requiring that one specifies exactly *which* properties change in an interaction (remember that Tabery thinks all activities are interactions) (Tabery 2004). The former corresponds to Tabery's argument that interactions need activities, and the latter to his argument that activities need interactions. In contrast, my modified conception does not prescribe use of an awkward neologism, but retains the more natural term 'activity'.

[24] As Tabery (2004) indicates, Machamer first suggested the term.

The problem with Tabery's interactivity thesis is that it synthesizes things of two very disparate ontological categories, and by doing so commits a category mistake.[25] As I state above in Sects. 2 and 3, interactions are mere terms of description while MDC's activities refer to producers of change that actually exist *in the world*. Thus, synthesizing interactions and activities is like combining *matter* and *force*—each belongs to a different and distinct category.[26] As such, they cannot be coherently synthesized. Furthermore, Tabery avows to be a monist (personal communication). But, as elementary arithmetic attests, 2 (MDC's dualism) +1 (Glennan's monism) does not equal 1 (Tabery's monism). Somewhere, a mistake has been made.

I want to talk briefly about how interventionism fits into my modified conception. I have already proposed a tentative definition of 'activity' based on Glennan's stipulation of 'interaction'. But how exactly should we understand the *characterization* of activities in my modified conception given in terms of "direct, invariant, change-relating generalizations"?[27] To begin, MDC write that "sometimes the regularities of activities can be described by laws. Sometimes they cannot" (MDC 2000, p. 7). (By "laws" MDC mean the traditional notion of universal, exceptionless generalizations.) For example, one cannot find a law according to which voltage-gated channels on an axon open when electrically stimulated, although they generally behave under such conditions in a regular manner.

On Woodward's view, philosophers ought to abandon talk of laws and adopt talk of invariance, which characterizes generalizations that hold over some domain of ideal interventions. The notion of *invariant* (rather than lawful) *generalizations* ensconces itself much more naturally within the biological sciences because entities often exhibit regular behavior only within a narrow range of perturbation. In other words, generalizations that describe their behavior involve existential quantification and allow for exceptions when, for example, background conditions preclude the relevant intra-mechanism parts from working properly.

Thus, by incorporating Woodward's manipulability theory, my modified conception allows one to characterize the activity of an intra-mechanism part by formulating invariant generalizations that describe that part's behavior. An advantage of this, as Craver (2007a) points out, pertains to the fact that scientists determine the relevance relations of negative causes *no differently* than they determine those of positive causes.[28]

My descriptivist interpretation of activities, in concert with Woodward's interventionism, is therefore congenial to a *genuinist* stance on causation.[29] Indeed, following Schaffer (2005), I champion genuinism as the correct view of causation, especially when considering disciplines like neuroscience, in which absences,

[25] I obviate this category mistake with my descriptivist interpretation of activities, which explicitly relieves activities of their ontological baggage.

[26] I thank an anonymous referee for this example.

[27] The term 'direct' here simply signifies that there is no intervening activity *on that mechanism level* between an entity and its behavior or two entities and their interaction.

[28] See Craver (2007a, p. 82), for a more detailed analysis.

[29] 'Genuinism' is Phil Dowe's coinage (Dowe 2004).

interferences, failures, non-occurrences, etc. are just as ubiquitous as positive causal factors [such as those involving the transfer of a conserved quantity (see Salmon 1994)]. For example, the retina involves a great deal of *lateral inhibition*, in which horizontal cells in the outer-plexiform layer inhibit photoreceptors surrounding those that are directly stimulated by light. This inhibition is just as essential to vision as the stimulation of photoreceptors in the "center" (rather than "surround").[30]

Furthermore, the mechanism by which photoreceptors respond to light involves quite a lot of negative causation. One need not adduce particular instances of neural mechanisms involving negative causes, though, to convince skeptics that such causation is important in neuroscience. Indeed, one only ask the neuroscientists themselves! Thus, if one wants to give an account of causation that is *scientifically plausible*, one must count instances of negative causes as genuine causes. This is precisely what my modified conception allows one to do.

6 Concluding Thoughts

In conclusion, the philosophy of mechanisms has emerged as a promising new area of philosophical research, offering novel approaches to the analysis of explanation and causation, as well as the composition of reality (Machamer 2002) and the nature of biological research (Darden 2006). The preliminary task of mechanicists is to formulate an adequate *conception of mechanisms*, which would prescribe norms for conceptualizing mechanisms.

In this paper, I argued that neither Glennan's nor MDC's conceptions satisfy the basic criteria of adequacy, but that my modified conception does. In sum, it does this by reinterpreting MDC's notion of activities such that activity verbs are explanatory by virtue of their descriptive content, rather than by reifying activities as the causal components of mechanisms—the "producers of change" (as in MDC's conception). Furthermore, following Glennan (2002), my modified conception characterizes intra-mechanism activities, understood according to my descriptivist interpretation, in Woodwardian terms of direct, invariant, change-relating generalizations. Together, these features allow my conception to obviate the problem of negative causation, an insoluble one for both Glennan's and MDC's accounts. Finally, my modified conception posits a monistic ontology of mechanisms, which thereby preserves metaphysical parsimony.

Acknowledgements I would like to thank an anonymous referee, Stuart Glennan, Christopher Cherniak, Mathias Frisch and Carl Craver for helpful feedback. Special thanks to Lindley Darden for many insightful comments on earlier drafts.

References

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

[30] Note that Craver also discusses inhibition as negative causation (Craver 2007a, pp. 6, 81).

Bechtel, W., & Richardson, R. (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.
Bogen, J. (2004). Analysing causality: The opposite of counterfactual is factual. *International Studies in the Philosophy of Science, 18*, 3–26.
Bogen, J. (2008). Causally productive activities. *Studies in History and Philosophy of Science, 39*, 112–123.
Craver, C. F. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.
Craver, C. F. (2007a). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
Craver, C. F. (2007b). *Activities and causal relevance, presented at Mechanisms and Causation*. University of Maryland, College Park.
Darden, L. (1991). *Theory change in science: Strategies from Mendelian genetics*. New York: Oxford University Press.
Darden, L. (2002). Strategies for discovering mechanisms: Schema instantiation, modular subassembly, forward/backward chaining. *Philosophy of Science, 69*, S354–S365.
Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: Cambridge University Press.
Darden, L., & Craver, C. (2002). Strategies in the interfield discovery of the mechanism of protein synthesis. *Studies in the History and Philosophy of the Biological and Biomedical Sciences, 33*, 1–28.
Dowe, P. (2004). Causes are physically connected to their effects: Why preventers and omissions are not causes. In C. Hitchcock (Ed.), *Contemporary debates in philosophy of science* (pp. 189–196). Oxford: Blackwell.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71.
Glennan, S. (1997). Capacities, universality and singularity. *Philosophy of Science, 64*, 605–626.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353.
Hume, D. (2006). *An enquiry concerning human understanding*. New York: Oxford University Press.
Machamer, P. (2002). Activities and causation, *Philosophy of Science Association* (PSA), 18th Biennial Meeting: Contributed Papers (Milwaukee, WI; 2002): PSA 2002 Workshops. http://philsci-archive.pitt.edu/archive/00000864/.
Machamer, P. (2004). Activities and causation: The metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*, 27–39.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Morris, W. E. (2008). David Hume, The Stanford Encyclopedia of Philosophy (Fall 2008 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/fall2008/entries/hume/.
Pietroski, P., & Rey, G. (1995). When other things aren't equal: Saving Ceteris Paribus laws from vacuity. *The British Journal for the Philosophy of Science, 46*, 81–110.
Purves, D., Augustine, G. J., Fitzpatrick, D., Hall, W. C., LaMantia, A. S., McNamara, J. O., et al. (2004). *Neuroscience*. Sunderland, MA: Sinauer Associates, Inc.
Quine, W. V., & Ullian, J. S. (1978). *The Web of Belief* (2nd ed.). McGraw-Hill Humanities/SocialSciences/Languages.
Rescher, N. (1996). *Process metaphysics: An introduction to process philosophy*. New-York: State University of New York Press.
Rescher, N. (2002). Process philosophy, The Stanford Encyclopedia of Philosophy (Summer 2002 Edition), E. N. Zalta (ed.), http://plato.stanford.edu/archives/sum2002/entries/process-philosophy/.
Rey, G. (1997). *Contemporary philosophy of mind: A contentiously classical approach*. Oxford: Blackwell.
Salmon, W. (1984a). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
Salmon, W. (1984b). Scientific explanation: Three basic conceptions. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science*, pp. 293–305.
Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science, 61*, 297–312.
Schaffer, J. (2003). Metaphysics of causation, The Stanford Encyclopedia of Philosophy (Winter 2007 Edition), E. N. Zalta (ed.), forthcoming http://plato.stanford.edu/archives/win2007/entries/causation-metaphysics/.
Schaffer, J. (2005). Contrastive causation. *Philosophical Review, 114*, 297–328.
Skipper, R. A., & Millstein, R. L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences, 36*, 327–347.

Tabery, J. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science, 71*, 1–15.

Thagard, P. (2006). What is a medical theory? In R. Paton & L. A. McNamara (Eds.), *Multidisciplinary approaches to theory in medicine* (pp. 47–62). Amsterdam: Elsevier.

Woodward, J. (2000). Explanation and invariance in the special sciences. *British Journal for the Philosophy of Science, 51*, 197–254.

Woodward, J. (2001). Law and explanation in biology: Invariance is the kind of stability that matters. *Philosophy of Science, 68*, 1–20.

Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science (Suppl.), 69*, S366–S377.