Euro Jnl Phil Sci (2012) 2:119–135
DOI 10.1007/s13194-011-0038-2

ORIGINAL PAPER IN METAPHYSICS OF SCIENCE

What is a mechanism? Thinking about mechanisms *across* the sciences

Phyllis McKay Illari · Jon Williamson

Received: 11 August 2010 / Accepted: 19 August 2011 / Published online: 13 September 2011

Abstract After a decade of intense debate about mechanisms, there is still no consensus characterization. In this paper we argue for a characterization that applies widely to mechanisms across the sciences. We examine and defend our disagreements with the major current contenders for characterizations of mechanisms. Ultimately, we indicate that the major contenders can all sign up to our characterization.

Keywords Mechanism · Explanation · MDC · Glennan · Bechtel · Astrophysical mechanism

1 Introduction

Since Bechtel and Richardson's 1993 book, there have been nearly two decades of debate on the right characterisation of a mechanism, intensifying since Machamer, Darden and Craver's (MDC's) controversial 2000 paper. The main contenders are:

> **Machamer, Darden and Craver**: 'Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions.' (Machamer et al. 2000 p3.)
> **Glennan**: 'A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.' (Glennan 2002b pS344.)
> **Bechtel and Abrahamsen**: 'A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena.' (Bechtel and Abrahamsen 2005 p423.)

P. M. Illari (✉) · J. Williamson
Philosophy Department, University of Kent, Canterbury Kent CT2 7NZ, UK
e-mail: p.mckay@kent.ac.uk

J. Williamson
e-mail: j.williamson@kent.ac.uk

After small changes of detail (see Bechtel and Richardson's original 1993, Glennan's original 1996; Machamer 2004; Craver 2007, and Glennan 2011), these broad characterisations remain in use by their original advocate(s), and many others.[1]

In this paper, we will defend a characterization that gives an understanding of what is common to mechanisms in *all* fields. We disagree with elements of all of the major characterizations above, and argue for:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.

This project is important for two reasons. First, it is important to the broad question of whether or not scientific method is disunified (see Glennan 2010). Different scientific disciplines share many methodological concerns, including causal explanation, causal inference and causal modelling, which commonly use mechanisms. It is our contention that we have produced a widely applicable understanding of mechanisms, that is of use in understanding what these different disciplines *share*, methodologically. This is *complementary* to the alternative project of describing what is *distinctive* about the kinds of mechanisms used in a particular domain. (See Steel; and Torres p240 for methodological disagreement.) Surface differences are methodologically important, but shouldn't be allowed to obscure what is common. Indeed, we cannot properly understand the differences without also seeing the similarities. We offer what is common to mechanisms, which different fields can flesh out with their distinctive methodological needs.

Second, these particular methodological debates and others *need* a consensus account of mechanisms. Philosophers and scientists are attempting to use mechanisms to illuminate causal explanation, inference and modelling, as well as the metaphysics of causality (see Glennan 1996; Steel 2008; Leuridan and Weber 2011; Broadbent 2011; Gillies 2011). These debates are impeded by lack of a consensus account, in spite of a great deal of consensus now existing within the mechanisms literature. To develop an understanding of the problems of causal explanation, inference and modelling that the sciences *share*, it is vital to understand what is common in the use of mechanisms across the sciences. The problems *shared* by different fields are just as important to recognise as the methodological differences (see Glennan 2005 p462). Many mechanistic explanations are built using components from multiple fields (see Craver 2007; Russo 2009; Illari and Williamson 2010). Debate on using mechanisms in causal inference includes both biomedical and social sciences (see for example Steel 2008; Gillies 2011). Such examples strongly indicate that mechanisms in general share a great deal. Finally, if there is no widely applicable account of mechanisms, there is no possibility of a widely applicable mechanistic approach to the metaphysics of causality, so our work is also of interest to that debate (Williamson 2011). We will assist all these debates by developing a consensus account of a mechanism that they can use.

We are interested in mechanisms themselves. As Craver claims, there is a sense of ontic explanation: mechanisms explain phenomena in the sense that their presence

[1] We lack space to discuss the work of everyone in the debate in detail, but we will also make some comments on the work of Tabery, Torres, and Woodward on the way through the paper.

produces the phenomenon (2007 pp27–8). But epistemic explanation is also important, as Bechtel claims, where the *description* of the mechanism explains the phenomenon (2008 p16). But both ontic and epistemic mechanistic explanation require real *mechanisms*. Bechtel and Abrahamsen write: 'mechanisms are real systems in nature' (2005 p424–5), and Bechtel agrees that epistemic explanation is parasitic on there being real mechanisms in the world to describe (private communication). So Bechtel and Craver can hold, with us, that examining mechanistic *explanation* tells you about mechanisms *themselves*, and so we will move freely between claims about mechanistic explanation and claims about mechanisms themselves.

Although our characterization is close to MDC's and Craver's, in the next section, S2, we explain why we do not include certain elements of the current characterizations. In S3 we defend our characterization. We will show that, correctly understood, it applies to the mechanisms that scientists discover and use in explanation and causal inference. Existing accounts of mechanisms have been developed in the light of the biomedical sciences (Machamer et al. 2000) and psychology (Craver 2007; Bechtel 2008). We will use astrophysical mechanisms to demonstrate the wide applicability of our account. In S4 we will take up the question of what is *not* a mechanism on this account. In S5 we conclude. While our *characterization* of mechanisms differs from those of other accounts, we see our *project* as consistent with those of the main contenders, and we briefly indicate why we think they should have no serious objections to our account.

2 How not to characterize mechanisms

With broad applicability in mind, we do not characterise a mechanism as a structure (Bechtel and Abrahamsen) or a system (Glennan). Unless read so weakly as to mean almost nothing, the idea of structure implies some level of inflexibility. This seems at odds with Bechtel's latest work (Bechtel and Wright 2009; Bechtel 2010, 2008; but compare Bechtel 2007 p275). A system is more dynamic and more flexible than a structure, but still implies a level of internal coherence that not all mechanisms show. As Darden notes (2006 p281), some mechanisms make their own entities as they go, such as the mechanism of protein synthesis where mRNA is made when needed and broken down afterwards. Further, many mechanisms are complex, but they can also be simple.[2] The mechanism of thermal dissociation of the diatomic iodine molecule in the vapour phase seems too simple to be called either a system or a structure. The stretching vibration just gets more and more energetic until there is enough energy to rupture the bond between two atoms, and they fly apart.

Unsurprisingly, astrophysical mechanisms are often relatively stable and structured. But violent sudden change from an existing structure or system to a different one is also possible, as with supernovae. Thus even for astrophysical mechanisms, it is best to avoid 'structure' or 'system' in the characterization of a mechanism.

[2] For this reason, we do not adopt Torres (2009, p247). At 'Mechanisms and Causality' conference, Kent, 2009, Glennan clearly withdrew 'complex' from his characterization.

Glennan is initially committed to all mechanisms being systems (2002b p128, p129; 2009a). In Glennan (2009b p323) he allows that there is no 'mechanism qua system' for a baseball breaking a window. In Glennan (2010, see especially pp260–1) he develops an account of 'ephemeral mechanisms', where the *configuration* of parts isn't stable, as it is in a system. In Glennan (2008, see especially p283) he calls for an account of a possible third kind, emergent mechanisms, for cases where phenomena produced by mechanisms depend on the properties of and relations between their parts, but standard mechanistic strategies such as functional localization are not very successful. We agree that these are all mechanisms, but are inclined to treat these differences as positions on a continuum, not differences in *kinds* of mechanism.

We also drop MDC's 'start or set-up' or 'finish or termination conditions'. Craver drops this without explanation (2007), while Darden (2006) and Machamer (2004) retain it. This element can be read very lightly, but it is better removed, because 'start' and 'finish' conditions are not even an aspect of all of our *mechanism descriptions*, far less of all *mechanisms*. They are pragmatic aspects of the descriptions we give of *some* mechanisms – but not all. Cell mechanisms such as the Krebs cycle are cyclical. They are continuing, having no real start or end. Bechtel (2009) notes this for other mechanisms. For continuing mechanisms, understanding that there is no tidy start or end is very important. Further, even some mechanisms that are neatly described in terms of start and finish conditions do not themselves have start and finish conditions. So while some mechanisms might have a natural descriptive starting point – we might start the description of the formation of stars with the gravitational accretion of dark matter in a halo – we should not enforce a start and end-point with a requirement in the characterization of a mechanism.

We follow MDC and Bechtel in not requiring modularity. Dynamical systems and systems biology explanations are precisely aimed at describing systems that are largely non-modular, and we do not wish to rule them out as mechanisms. It may appear that we disagree with Woodward. However, on closer examination, Woodward is talking only about *representations* of mechanisms: '(MECH) a necessary condition for a *representation to be an acceptable model of a mechanism* is that the representation (i) describe an organized or structured set of parts or components, where (ii) the behavior of each component is described by a generalization that is invariant under interventions, and where (iii) the generalizations governing each component are also independently changeable, and where (iv) the representation allows us to see how, in virtue of (i), (ii) and (iii), the overall output of the mechanism will vary under manipulation of the input to each component and changes in the components themselves.' (Woodward 2002 S375, emphasis added. Compare Darden 2006 p279.) But Woodward is clear here that he is concerned with representations or models, not mechanisms themselves. We agree that our representations or models of mechanisms should be modular as far as possible. Such a representation will certainly make prediction and intervention easier. But where this is not possible, a non-modular representation will have to do. Neither all mechanisms themselves, nor all mechanism descriptions, will be modular. Since Woodward's primary concern is representations, not mechanisms themselves, we will put his views aside for the rest of the paper. Thus, we do not use Woodward's ideas in the characterization of a mechanism itself.

We will move on now to defending our positive characterization of a mechanism. This is our characterization of the consensus elements of preceding accounts. Here, disagreements are more subtle, but they are important if the characterization is to be widely applicable and so useful to other debates on method or metaphysics.

3 Our characterization of mechanisms

All mechanistic explanations begin with (a) the identification of a phenomenon or some phenomena to be explained, (b) proceed by decomposition into the entities and activities relevant to the phenomenon, and (c) give the organization of entities and activities by which they produce the phenomenon. (See Darden 2006; Bechtel and Abrahamsen 2008.) Mechanism discovery is often messy and iterative, but always involves finding out about these three elements.

This is widely known, so all mechanisms share the three elements found in the process of mechanism discovery. Even astrophysical mechanisms are grouped by the phenomena they produce. Scientists aim to give a detailed account of *how* the phenomenon is produced by entities and activities. Entities include both massive bodies such as stars and galaxies, and fundamental particles such as quarks, photons, neutrons and neutrinos; while activities tend to involve movement and energy changes. Organization is vital: threshold effects are common, and feedback effects, often associated with biological mechanisms, are not uncommon. Background theory, particularly General Relativity, is important to astrophysical mechanisms in a way not paralleled by all mechanisms, because relevant organization can include details of background space-time geometry.

Our favoured characterisation captures the core consensus at the heart of the views of the main contenders:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.

In the following subsections, we take each of the three elements here and argue for them: responsible for the phenomenon (S3a), entities and activities (S3b), and organization (S3c). We have explored entities, activities and organization in the context of protein synthesis and natural selection in more detail elsewhere (Illari and Williamson 2010).

3.1 Responsible for the phenomenon

There are three reasons why we follow Bechtel and Abrahamsen (2005 p422) in saying mechanisms are ‘responsible for a phenomenon’. The first reason is the importance of the phenomenon for mechanistic explanation. Mechanistic explanation succeeds when the mechanism discovered and described is the mechanism *responsible for the phenomenon*. If no unified mechanism can be found for that phenomenon, the phenomenon is redescribed to make it susceptible of mechanistic explanation – what Bechtel and Richardson call ‘reconstituting the phenomenon’ (1993). This is to say that mechanisms are functionally individuated by their phenomena.[3] However, we avoid Bechtel and Abrahamsen's 'performing a function' in our characterization. In wider philosophical and scientific debate, 'function' is a loaded concept, usually involving deliberate design or natural selection, while the function of a mechanism requires only something like 'characteristic activity'. We therefore avoid using 'function' explicitly in the characterization of the mechanism.

This point is important in the case of astrophysical mechanisms, which involve neither deliberate design, nor natural selection, but do have characteristic activities. Even in the absence of natural selection or deliberate design, spectacular phenomena such as supernovae are typed by the mechanisms that produce them.[4] In a supernova of Type II the star explodes but leaves a collapsed black hole, neutron star or white dwarf behind. The core has little nuclear material left, and is supported by electron degeneracy pressure (when compressed and cooled, the velocity of all electrons can only fall so low because two electrons can't occupy the same quantum state). But the core accumulates mass from the shell. If it never reaches the Chandrasekhar mass, it will collapse to a white dwarf. But when the core mass is larger than the Chandrasekhar mass, electron degeneracy pressure is not enough, and it collapses further. When *neutron* degeneracy pressure starts the bounce, many neutrinos escape suddenly, carrying away an enormous amount of energy, leaving a neutron star behind but blowing away the rest of the mass of the star. Supernovae of type I are different – they are giant nuclear explosions. In a supernova of Type Ia (characterized by absence of hydrogen lines in their spectra), the star explodes completely, leaving nothing behind. The star still has nuclear material, and during collapse increasing density and pressure rapidly increases nuclear reactions, which release energy. This stops collapse well before neutron star density, blowing the star completely apart. Even here we redescribe and regroup phenomena, paying more attention to some differences than others, when we discover that there is more than one mechanism for supernovae.

'Responsible for a phenomenon' expresses this. Secondly, it captures the *diversity* of things that mechanisms do. Mechanisms carry out tasks, such as regulation or control, and exhibit behaviours, such as growth. They also maintain stable states. Homeostatic mechanisms, such as those that maintain human body temperature at 37°, do this. Such a state might even be a standing capacity of a system. For example, many cells have the capacity to metabolise lactose, although they do not do so unless glucose is unavailable. At a higher degree of abstraction, the metabolic mechanism is responsible for more than one phenomenon: metabolising glucose normally, and metabolising lactose in the absence of glucose. There is no significant disagreement on this diversity (Darden 2006 p273, 2008 p959; Glennan 2002a p126–7).

[3] At least partially. There seem to be *other* ways to individuate mechanisms that produce the same phenomenon, such as in terms of the entities or activities involved, and an examination of whether such ways can always be explained away in terms of functional individuation is a complex issue we reserve for further work. At 'Mechanisms and Causality' conference, Kent 2009, both Darden and Craver called the functional individuation of mechanisms 'Glennan's Law', as he was the first to recognise this (see for example his 1996).

[4] Mechanisms are individuated by their phenomena, and phenomena are also individuated by their mechanisms. This is not circular, because it happens iteratively over time. At the beginning, a mechanism is not needed to individuate a phenomenon, but the characterization of the phenomenon may be further refined when a mechanism or mechanisms are discovered. See Darden 2008 p960.

Thirdly, ‘responsibility’ implies something counterfactual. The phenomenon can be something actual, or something modal – such as the capacity of a cell to metabolise lactose, even if lactose is never encountered. See Glennan (1997) for a similar view of capacities. However, the mechanism does not *determine* the phenomenon, because some mechanisms may be indeterministic. Nor should a characterization of mechanisms require that they produce ‘*regular changes*’ as MDC do, but Machamer (2004 p37, footnote 1) drops. Compare Darden (2008 p964) and Glennan (2010 p257). Some mechanisms, such as homeostatic mechanisms, might not produce change at all.[5] They may or may not be regular. To give Craver's example, in the mechanism of neurotransmitter release, only 10–20% of action potentials eventuate in release events. And release events can occur without action potentials (Craver 2007 p26). But dropping explicit reference to regularity does not imply that mechanisms in general do not have to exhibit some form of regularity or stability. *Some* far weaker form of regularity or stability is already present in the idea of mechanisms being responsible for the phenomenon. Our formulation captures the importance, diversity and various forms of stability of what mechanisms do.

3.2 Entities and activities

There is consensus that mechanistic explanation involves *decomposition* (see particularly Bechtel and Richardson), and mechanisms have two distinct *kinds* of constituents. We have ‘entities’, ‘parts’ and ‘component parts’ used for the bits and pieces of the mechanism, and ‘activities’, ‘interactions’ and ‘component operations’ for what those bits and pieces do. Astrophysical mechanisms have both entities and activities: ‘An important mechanism for producing X-rays from Solar System objects is charge exchange, which occurs when a highly ionized atom in the solar wind collides with a neutral atom (gas or solid) and captures an electron, usually in an excited state. As the ion relaxes, it radiates an X-ray characteristic of the wind ion. Lines produced by charge exchange with solar wind ions such as C V, C VI, O VII, O VIII and Ne IX have all been detected with Chandra and XMM-Newton [new space observatories]…’. (Santos-Lleo et al. 2009. p998.) Putting this together with mechanisms for supernovae above, entities include: electron, proton, neutron, neutrino, star, neutron star, white dwarf, black hole, core, gas, x-ray, ionised atom, solar wind, neutral atom. Activities include: charge exchange, colliding, relaxing, radiating, collapse, bounce, heating, electron capture, the nuclear fusion that creates heavier elements in stars, and so on.

For wide applicability, care is needed in understanding entities and activities. Fascinatingly, astrophysical mechanisms deal simultaneously with the vanishingly small and the staggeringly enormous. What happens in a supernova depends on properties of the massive, such as whether the star's core reaches the Chandrasekhar mass or not – which is approximately 1.2–1.4 solar masses. On the other hand, it is electron degeneracy pressure which supports a white dwarf, and this depends on the fact that electrons are fermions, i.e. they obey Pauli's Exclusion Principle, which means that there are limits on the minimum energy that more than two electrons in

[5] We reject Tabery's ‘interactivity’ because it also requires change (Tabery 2004 p12). But see Tabery (2009) on using mechanisms to explain *difference*, rather than similarity.

the same place can have. The end state of a star depends on the interplay of these very different kinds of factors, so there can be no a priori restriction according to size on the entities and activities of a mechanism. Further, mechanistic explanation might not always be in terms of *smaller* parts. Darden provides a good example: 'finding the mechanism of segregation of genes did not require decomposing genes into their parts but required finding the wholes, the chromosomes, on which the parts, the genes, ride.' (Darden 2006 p109, see also Darden 2008 p961.) Mechanistic explanation is not *always* about the little explaining the big. Finally, the parts of mechanisms vary a great deal in their robustness. Some entities remain comparatively unchanged over time, but others are more transient, such as the mRNA that is made from DNA, used as a template to make a protein, and then broken down again straight away. Activities can also be local and fragile, such as the mutation or recombination that creates the diversity of strains of HIV that makes it so difficult to eradicate. Glennan seems committed to a high degree of robustness of parts in earlier work (2002b, 2009a) – although he notes that the interactions of parts are 'not exceptionless' – but has relaxed this somewhat now (2010).

MDC have metaphysical arguments for entities and activities. Here, we put these aside to focus on descriptive reasons for preferring a particular characterization of the components of a mechanism. We prefer MDC's language of activities and entities for two main reasons: it offers a powerful resistance to *entity-bias*, and it allows *variability in the arity of the relation* between entities. We take these points in turn.

Many approaches to scientific ontology give entities priority, treating what entities *do* as either reducible to entities themselves, or metaphysically dubious. But descriptively, activities and entities are *equally* important to mechanisms: *neither* has priority. MDC write: 'There are kinds of changing just as there are kinds of entities. These different kinds are recognized by science and are basic to the ways that things work.' (Machamer et al. 2000 p5.) Machamer adds: 'Activities can be abstracted and referred to and identified independently of any particular entity, and sometimes even without reference to any entity at all.' (Machamer 2004 p30. See also Darden 2006 p277.) A bunch of entities engaging in a certain set of activities will produce something different from the same bunch of entities engaging in another set of activities. A buyer and seller haggling over the price may lead to a sale. The same two people chatting about the weather will not. Further, although entities and activities are always equally important in that they must both be present to produce the phenomenon, in explaining different kinds of phenomena entities are sometimes more interesting than the activities, and vice versa. In protein synthesis, entities are very different from each other and their detailed structure matters a great deal. But in many dynamical systems and systems biology explanations, the entities are relatively similar to each other and the activities are vital to produce the phenomenon.

This is consistent with Bechtel's and Glennan's considered views (Glennan 2009b p321), but the rhetorical impact of the language matters for scientists and philosophers elsewhere using an account of mechanisms in other debates. MDC's entities and activities offer the strongest rhetorical resistance to a default entity-bias.

Our second reason is that variability in the *arity* of the relation between entities is more important than has been recognised, and is nicely captured by MDC's language. Consider the alternatives: capacities are unary (1-ary) relations since a capacity attaches to an entity, although one entity can have many capacities (note Darden 2008 p963). Glennan's 'interaction' implies a relation between at least two entities, so interactions are binary (2-ary) at least.[6] Bechtel and Abrahamsen write: 'Each component operation involves at least one component part' (2005 p424), which seems to allow either unary, binary, 3-ary and so on. The mapping of entities to activities can be unary, as in a bond *breaking*, involving no other entity; binary, as in a promoter *binding* to a strand of DNA; but it can also be 3-ary, 4-ary and so on (See Darden 2008 p964). The activity of *transcription* involves DNA, the newly created mRNA, and various regulation and control enzymes, while more highly abstract activities such as *equilibrating*, or *osmosis* (Darden 2006 p277) may involve very many entities, of the same or different kinds, or be such that it is hard to decide on any very clearly defined entity that engages in the activity.

Bechtel (2008) examines extensively the importance of mapping entities to activities (his component parts and component operations) in mechanism discovery, pointing out that it is often this mapping that allows us to *identify* the working parts of a mechanism. So we had better get the arity of the relation right. But Bechtel ties operations closely to parts: 'We use the term operation rather than activity because we want to draw attention to the involvement of parts' (Bechtel and Abrahamsen 2005 p423, footnote 5). The arity of the relation between entities allowed by activities is unrestricted, covering all this. This is the best descriptive reason to favour entities and activities.

In summary, mechanisms have two kinds of constituents. We prefer 'entities' and 'activities' because these terms offer rhetorical advantages for avoiding entity-bias, and there is no limit to the arity of activities. Entities can be of widely varying sizes, in some cases the big is used to explain the small, and some mechanisms involve comparatively fragile entities and activities.

3.3 Organization

Organization is the least controversial element in any characterization of mechanisms, present in the characterizations of MDC and Bechtel and Abrahamsen, and discussed explicitly by Glennan elsewhere (see 2005, 2002a). We think it worth the emphasis of putting it in the characterization, but consider Bechtel and Abrahamsen's 'orchestrated functioning' too strong. It suggests a tightly integrated form of organization that exists in highly evolved or designed systems, but not everywhere.

How to *understand* organization is not much discussed, and is far from trivial. What is organization so that it can reasonably be regarded as an element of *all* mechanisms? Here, we examine this, and argue that organization is *not* confined to complex biological mechanisms, by showing its importance to astrophysical mechanisms. These exhibit complex forms of organization requiring investigation by numerical simulation, such as homeostasis, equilibrium and feedback.

Organization is the final element in the production of the phenomenon. The same entities and activities organized differently will produce something different. A group of organisms engaged in feeding, mating and dying will do something different if they are subject to a common selection pressure – a new predator, or bout

[6] As Tabery 2004 notes. We thank Glennan for pressing us on this point.

of cold weather – than if they are not. Most generally, organization is whatever relations between the entities and activities discovered produce the phenomenon of interest: when activities and entities each do something and do something together to produce the phenomenon.[7]

In mechanistic explanation, organization is analogous to initial conditions in laws-based explanation. Laws and the entities they govern explain nothing until initial conditions are specified: Newton's laws do not tell us the movements of the planets until their initial positions and velocities are specified. In the mechanistic approach organization gives the *ongoing conditions* that allow the entities and activities to produce the phenomenon. 'Ongoing' is important. Initial conditions for laws matter only at the beginning, while organization matters *throughout* the operation of a mechanism. Further, organization is *not* independent of the activities and entities and ongoing operation of a mechanism. Organization might affect which activities and entities are involved, while the operation of a mechanism might alter the organizational structure. Evolution of a group of organisms subject to a common selection pressure might alter how widely dispersed those organisms *need* to be to be subject to that common selection pressure.

This approach implies, correctly, that it is an empirical question *what* forms of organization are important for *particular domains*, so that the only other informative thing that can be said about organization is to discuss examples. Organization comes in many forms, more or less important for different kinds of mechanism. Spatial and temporal organization is vital to such cases as protein synthesis. (Darden 2006; Craver 2007). But other forms of organization can be instantiated by spatiotemporally located mechanisms. Complex forms of organization such as homeostasis, equilibrium, feedback and self-organization are vital for the production of the phenomena studied by complex and dynamical systems. (See Bechtel 2006 p33, p39; Mitchell 2003; and possibly Glennan 2008.) Quantitative description of dynamical organization is often vital. For example, in simulating supernovae, mass is standardly being *lost* from the star while mass is *accumulated* in the star core. Quantitative simulation over time is needed to see whether the Chandrasekhar mass is reached. In this way we allow organization to capture necessary elements of what Bechtel calls 'dynamical mechanistic explanation'. (See Bechtel 2008; Bechtel and Abrahamsen 2009, 2010). Each of these forms of organization also lies on a spectrum from less organized to increasingly organized. Unsurprisingly, then, organization in its most general form – when activities and entities each do something and do something together to produce the phenomenon – itself comes in a (multidimensional) spectrum of increasingly complex organization. Whichever form of organization is most important to the production of a particular phenomenon depends on the empirical world. Our world seems to involve different forms of organization, more or less complex, in different cases. In the simplest cases organization might be simple or trivial, but it is still present.

Use of *numerical simulation* is a good indicator of complexity of organization, and simulations are a standard tool for discovering astrophysical mechanisms. They often reveal complex forms of organization usually associated with biological mechanisms such as *feedback*. Simulation of how the first stars formed tend to

[7] We have compared organization in natural selection, and in protein synthesis in Illari and Williamson (2010).

suggest they formed on their own, which leads to the question: how did galaxies form? Further simulations suggest: 'Some of the feedback processes described above that affect the formation of individual stars also influence primordial star formation on large scales. The enormous fluxes of ionizing radiation and H_2-dissociating Lyman–Werner radiation emitted by massive population III stars dramatically influence their surroundings, heating and ionizing the gas within a few kiloparsecs of the progenitor and destroying the H_2 within a somewhat larger region. Moreover, the Lyman–Werner radiation emitted by the first stars could propagate across cosmological distances, allowing the buildup of a pervasive Lyman–Werner background radiation field. The effect of radiation from the first stars on their local surroundings has important implications for the numbers and types of population III stars that form. The photoheating of gas in the minihaloes hosting population III.1 stars drives strong outflows, lowering the density of the gas in the minihaloes and delaying subsequent star formation by up to 100 Myr … . Furthermore, neighbouring minihaloes may be photoevaporated, delaying star formation in such systems as well. The photodissociation of molecules by Lyman–Werner photons emitted from local star-forming regions will, in general, act to delay star formation by destroying the main coolants that allow the gas to collapse and form stars.' (Bromm et al. 2009 p51.)

Successful simulations are often very difficult: 'The simulations, starting from cosmological initial conditions, are just now approaching the resolution and physical realism required to investigate whether atomic cooling haloes fulfil the criteria for a first galaxy as defined above. Quite generically, in such models, the first generation of stars forms before galaxies do, and feedback effects from the first stars are expected to play a key role in determining the initial conditions for the formation of the first galaxies.' (Bromm et al. 2009 p52.) Astrophysicists want to reproduce phenomena using *physically realistic parameters*, and only then do they think they have an empirically significant result. Investigation of organization by means of simulation is not the sole preserve of the life sciences.

We have now defended our characterization of a mechanism, argued for its wide application, including to the case of astrophysical mechanisms. We have indicated where we disagree with the main contenders while emphasizing that there is a core of agreement which we capture. Very different scientific work in different fields aims to find and describe the entities and activities of their domain, their organization, and the phenomena they are responsible for. This discovery process can be complicated and iterative. It takes serious empirical work to correctly delimit the phenomena, and that description determines what activities, entities and organization will be looked for; while what activities, entities and organization are found affect the description of the phenomena.

We will now show that our characterisation of mechanism is not so broad that it captures non-mechanisms.

4 Borderline cases

It is important that not everything counts as a mechanism. In this section, we examine some things produced in this process of discovery that are sometimes called mechanisms – perhaps erroneously – to further illuminate our account of mechanisms.

4.1 Case 1: Description of mechanism is too partial

Sometimes we have a scientific advance, but the *description* of the mechanism for the phenomenon is still partial. Consider the various possible forms of memory that have at some point been phenomenally dissociated: long term versus short term memory, working memory, episodic versus semantic memory, and non-explicit memory including various forms of priming. There may be separate mechanisms producing these phenomena, but we are not yet even in a position to guess how many mechanisms there are. (See Bechtel 2008 for extensive discussion.) We have a better description of the phenomenon to be explained, and only finding the underlying entities will show that there really are separate mechanisms. *Before* this point, we have little more than a better description of the *phenomenon* to be explained.

Mechanism discovery is gradual, so there will be no sharp line between partial and full descriptions of mechanisms. The crucial point is where scientists have good reason to suppose they have got hold of the actual mechanism operating. Before that, the description might be so partial that it does not pick out a *mechanism*, and the explanation might not succeed.

4.2 Case 2: Entities without activities – Darden's stopped clock

Darden writes: 'The MDC characterization of mechanism points to its *operation*. Although someone (perhaps Glennan 1996) might call a stopped clock, for example, a mechanism, I would not. It is a machine, not a mechanism. The MDC characterization views mechanisms as inherently active. In the stopped clock, the entities are in place but not operating, not engaging in time-keeping activities. When appropriate set-up conditions obtain (e.g., winding a spring, installing a battery), then the clock mechanism may operate.' (Darden 2006 p280–1.)

Recall that nothing is a mechanism *tout court* – mechanisms are mechanisms for phenomena. A stopped clock is no longer a mechanism for telling the time, but it might still be a mechanism for something else – for recording a race time. Recall also that for Darden, as for Machamer and Craver, activities must produce *change*. The stopped clock produces no change. But we have argued that some activities and mechanisms, such as homeostatic ones, exist to prevent change. So the stopped clock, and similar cases such as chimneys, or pillars supporting roofs, are candidate mechanisms for maintaining stability of some kind.

However, they still present a puzzle: it seems they must either be mechanisms without activities, or non-mechanisms due to the lack of activities. The normal explanation for a pillar supporting a roof involves only its material, spatio-temporal location and forces. This seems to involve organization and no activities. But this is too quick, as there is no sharp line between activities and organization. In one explanation, a high-level activity such as equilibrating might be the activity of a particular group of entities, while in another it is treated as the organization of the system. Ultimately in such cases there is no sharp answer to the question of whether these are cases of mechanisms without activities, and there is no useful purpose in legislating an answer to the question that could constrain empirical research.

4.3 Case 3: Activities without entities

In psychology in particular, capacities like memory are often given a purely functional description. There are indefinitely many ways the human brain could divide up the task of remembering things. There may be one part of the brain that remembers events, including that event from the point of view of all the sensory modalities, and facts. Alternatively, there could be different parts of the brain that remember these different kinds of things. For example, perhaps remembering what an event *looked like* is a sub-task of the *visual system*, rather than of a central memory area.

As we have said, finding the parts of the brain that perform these tasks – the working entities – is what persuades us that we have the *right* division of the overall task into sub-tasks. Where this is not possible, we are left with a purely functional explanation, which seems to consist of postulated activities, without entities. This is not a mechanism, although it may be a step on the way to a mechanistic explanation (Craver 2006). See also Bechtel (2008).

4.4 Case 4: No organization

There may appear to be no mechanism if there is no apparent organization. In the kinetic theory of gases, which explains both Boyle's Law and Charles' Law, molecules behave on average *randomly*.[8] But in our understanding of organization as when activities and entities each do something and do something together to produce the phenomenon, *whatever* relations amongst the activities and entities produces the phenomenon is the relevant organization. If the molecules behaving randomly on average produces the phenomenon, that is the kind of organization present in that mechanism, however trivial it appears.

This is not the same as the idealization case. If a false assumption is made of average random behaviour to model a system, that might – or might not – render the model no longer a model of a mechanism, as we have said above. But if the assumption is not false, a mechanism is being described.

4.5 Case 5: Nothing concrete

Mathematicians sometimes speak of 'mechanism', for a technique or schematic method. These techniques are normally mechanisms for generating derivations or mathematical entities or structures. For example, forcing is a mechanism for deriving the independence of the continuum hypothesis, and Foreman and Magidor (1995, p55) write of 'the mechanism typically used to show presaturation'.

The 'mechanisms' here are purely abstract. They are not causes, and cannot be used in causal inference or explanation. However, these things are used in explanation, prediction and control in the particular way appropriate to the abstract

[8] We thank Erik Weber for suggesting this example.

realm. There is an analogous form of explanation in the decomposition to parts, and the understanding of how parts together produce the overall derivation, entity or structure. They might also be used to predict a change in the overall result from changing a part – a prediction that couldn't be made before the decomposition.

These strong analogies render using the word 'mechanism' reasonable. To decide further whether these things count as mechanisms on our account depends on metaphysical issues we do not address here. Do entities and activities have to be concrete? If so, then these are not mechanisms, on most understandings of mathematical entities. However, even on this view a mathematical Platonist might make a case for these being real mechanisms. For our purposes it suffices to note that even in that case, there are clearly no *causal* mechanisms here.

4.6 Case 6: Too much idealization

As Glennan (2005) and Craver (2006) discusses extensively, *models* of mechanisms are built using assumptions. These are necessary for enough simplification to build a model. Sometimes these assumptions are radically false. For example, in the social sciences it is not uncommon to assume non-communication among people or groups – an assumption of no organization. In economics, it is standard to assume rationality. Many models in physics use equilibrium assumptions or no-friction assumptions. Often, these claims are trivial, merely allowing serious quantitative modelling of a genuine worldly phenomenon. But once there is too much idealization, these are no longer accurate models of mechanisms. They are too distant from the system they describe, and their parameters no longer have plausible physical interpretations. They might be useful predictive tools, or important explanatory work on the road to mechanism discovery. Such models are often of further use as accurate descriptions of phenomena to be explained. But scientists using such models are, as above, not yet in a position to know whether they have got hold of the actual mechanism.

The level of idealization versus the level of accurate description is a matter of degree, so there is no particular point where such models cease to be accurate descriptions of mechanisms. The crucial point is whether they accurately describe anything worldly, whether their parameters have reasonable physical values (see Bechtel and Wright). There may still be mechanisms in such cases – but such models have not yet described them. This extends to many models in science.

5 Conclusion

We have argued for our characterization of mechanisms:

> A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon.

We have examined the various elements – responsibility for the phenomenon; entities and activities; organization – in some detail, showing how they apply to various fields, particularly astrophysical mechanisms. However messy the process of mechanism discovery is, and however important the different challenges faced by different fields, this characterization lets us see how these elements of mechanisms contribute to a project that shares a great deal across the sciences. We hope that our account will be useful to ongoing debates on causal inference and causal modelling.

We believe our account best captures a consensus emerging in the mechanisms literature by applying very widely to mechanisms while addressing the primary concerns of the main contenders in the debate.

All the main contenders agree on the functional individuation of mechanisms – Glennan's Law. Indeed, many have worked on the implications of this (Craver 2007 pp6 ff; Darden 2006 p42, pp289-90; Bechtel 2006 p28). We have used this to frame our account, spelling out further implications, and there is no obvious reason for the main contenders to object. We have already explained that our use of MDC's entity-activity language is not at serious odds with Glennan's or Bechtel's considered views. Finally, there is little extended discussion of organization, so it is possible for the main contenders to regard our views on organization as a development of theirs.

Bechtel, Craver, Darden and Machamer do not *aim* for a widely applicable account of a mechanism, but they should have no objection to that aim. Craver and Bechtel are currently extending the applicability of mechanisms, at least to psychology and neuroscience. They have no reason to object to dropping those elements of their own characterizations that narrow their applicability.

Glennan *does* aim for a widely applicable account. He wishes to use an account of mechanisms to give an account of causation, so his account of mechanisms *must* apply anywhere there is causation. But we have argued that Glennan's wish in earlier work for stability of mechanisms and mechanism parts, and his definition of mechanisms as 'complex systems' narrow the applicability of his account. This creates serious tension in Glennan's work. Glennan most of all has excellent reason to alter these elements of his own characterization in favour of an account like ours, which explains why he is now moving in that direction (2010, 2009b).

In conclusion, we have offered a characterization of mechanisms that is widely applicable across the sciences and captures the emerging consensus on mechanisms. It is fit for use as a framework for ongoing work on causal explanation, inference and modelling.

Acknowledgements We wish to thank the Leverhulme Trust for supporting this research. We are also indebted to colleagues at Kent and in the Causality in the Sciences network for discussion of many of these issues. The work has been significantly improved due to detailed comments from Stuart Glennan, Federica Russo and two anonymous referees. Remaining errors are, of course, our own.

References

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

Bechtel, W. (2007). 'Biological mechanisms: Organized to maintain autonomy'. In Boogerd, Bruggeman, Hofmeyr and Westerhoff (Eds.), *Systems biology*, 2007, Elsevier.

Bechtel, W. (2008). *Mental mechanisms*. Oxford: Routledge.

Bechtel, W. (2009). Looking down, around, and up: mechanistic explanation in psychology. *Philosophical Psychology, 22*, 543–564.

Bechtel, W. (2010). The downs and ups of mechanistic research: circadian rhythm research as an exemplar. *Erkenntnis, 73*(3), 313–328.
Bechtel, W., & Abrahamsen, A. (2005). Explanation: a mechanist alternative. *Studies in the History and Philosophy of the Biological and Biomedical Sciences, 36*, 421–441.
Bechtel, W., & Abrahamsen, A. (2008). From reduction back to higher levels. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 559–564). Austin: Cognitive Science Society.
Bechtel, W., & Abrahamsen, A. (2009). Decomposing, recomposing, and situating circadian mechanisms: Three tasks in developing mechanistic explanations. In H. Leitgeb & A. Hieke (Eds.), *Reduction and elimination in philosophy of mind and philosophy of neuroscience* (pp. 173–186). Frankfurt: Ontos.
Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science, Part A, 1*, 321–333.
Bechtel, W., & Richardson, R. (1993). *Discovering complexity*. Princeton: Princeton University Press.
Bechtel, W., & Wright, C. (2009). What is psychological explanation? In P. Calvo & J. Symons (Eds.), *The Routledge companion to philosophy of psychology* (pp. 113–130). London: Routledge.
Broadbent, A. (2011). Inferring causation in epidemiology: Mechanisms, black boxes, and contrasts. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 45–69). Oxford: OUP.
Bromm, V., Yoshida, N., Hernquist, L., & McKee, C. F. (2009). 'The formation of the first stars and galaxies'. *Nature, 459*, 7 May 2009.
Craver, C. (2006). When mechanistic models explain. *Synthese, 153*, 355–376.
Craver, C. (2007). *Explaining the brain*. Oxford: Clarendon.
Darden, L. (2006). *Reasoning in biological discoveries*. Cambridge: CUP.
Darden, L. (2008). Thinking again about biological mechanisms. *Philosophy of Science, 75*(5), 958–969.
Foreman, M., & Magidor, M. (1995). Large cardinals and definable counterexamples to the continuum hypothesis. *Annals of Pure and Applied Logic, 76*(1), 47–97.
Gillies, D. (2011). The Russo-Williamson thesis and the question of whether smoking causes heart disease. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 110–125). Oxford: OUP.
Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*(1), 49–71.
Glennan, S. (1997). Capacities, universality, and singularity. *Philosophy of Science, 64*(4), 605–626.
Glennan, S. (2002a). Contextual unanimity and the units of selection problem. *Philosophy of Science, 69*, 118–137.
Glennan, S. (2002b). Rethinking mechanistic explanation. *Philosophy of Science, 69*, S342–S353.
Glennan, S. (2005). Modeling mechanisms. *Studies in the History and Philosophy of Biology and Biomedical Sciences, 36*, 443–464.
Glennan, S. (2008). Mechanisms. In S. Psillos & M. Curd (Eds.), *Routledge companion to the philosophy of science*. Oxford: Routledge.
Glennan, S. (2009a). Productivity, relevance and natural selection. *Biology and Philosophy, 24*, 325–339.
Glennan, S. (2009b). Mechanisms. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The Oxford Handbook of causation* (pp. 315–325). Oxford: OUP.
Glennan, S. (2010). Ephemeral mechanisms and historical explanation. *Erkenntnis, 72*, 251–266.
Glennan, S. (2011). Singular and general causal relations: A mechanist perspective. In Illari, Russo, & Williamson (Eds.), *Causality in the sciences* (pp. 789–817). Oxford: OUP.
Illari, P. M., & Williamson, J. (2010). Function and organization: comparing the mechanisms of protein synthesis and natural selection. *Studies in the History and Philosophy of the Biological and Biomedical Sciences, 41*, 279–291.
Leuridan, B., & Weber, E. (2011). The IARC and mechanistic evidence. In P. M. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 91–109). Oxford: OUP.
Machamer, P. (2004). Activities and causation: the metaphysics and epistemology of mechanisms. *International Studies in the Philosophy of Science, 18*(1), 27–39.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*, 1–25.
Mitchell, S. D. (2003). *Biological complexity and integrative pluralism*. Cambridge: CUP.
Russo, F. (2009). *Causality and causal modelling in the social sciences: Measuring variations*. New York: Springer.
Santos-Lleo, M., Schartel, N., Tenenbaum, H., Tucker, W., & Weisskopf, M. C. (2009). 'The first decade of science with Chandra and XMM-Newton'. *Nature*, *462*, 24/31, December 2009.
Steel, D. (2008). *Across the boundaries: Extrapolation in biology and social science*. OUP, 2008.

Tabery, J. (2004). Synthesizing activities and interactions in the concept of a mechanism. *Philosophy of Science, 71*, 1–15.
Tabery, J. (2009). Difference mechanisms: explaining variation with mechanisms. *Biology and Philosophy, 24*, 645–664.
Torres, P. J. (2009). A modified conception of mechanisms. *Erkenntnis, 71*(2), 233–251.
Williamson, J. (2011). Mechanistic theories of causality. *Philosophy Compass, 6*(6), 421–447.
Woodward, J. (2002). What is a mechanism? A counterfactual account. *Philosophy of Science, 69*, S366–S377.