Mechanisms and Resources*

Colin Klein
Macquarie University
cvklein@gmail.com

1 Introduction

Mechanistic accounts of explanation are popular (Machamer et al., 2000; Craver, 2007; Bechtel, 2008). They are also controversial. Most of the debate has been over whether mechanistic accounts are sufficiently *comprehensive* — that is, whether they cover all the explanations given in a particular domain. Everyone agrees that *some* explanations (especially in neuroscience and molecular biology) are mechanistic. Explanations in chemistry, or physics, or evolutionary biology are less obviously mechanistic, however. Even within neuroscience there are explanations that don't look terribly mechanistic: those which appeal to general network properties, or to dynamic laws, or to computational processes, all present special challenges (Stepp et al., 2011; Piccinini and Craver, 2011; Kaplan and Craver, 2011; Levy and Bechtel, 2013). In the background, there is also a general worry about whether mechanists properly distinguish models of mechanisms from descriptions of those models (Klein, 2012; Craver and Darden, 2013).

All of these are interesting, big-picture worries. I want to put them to one side and focus on a second kind of question. That's the issue of whether existing mechanistic accounts are *complete* – that is, whether they say all there is to say about explanations in the obviously mechanistic realms. Take some paradigmatic mechanistic explanations: explaining the action potential, or how an engine works, or how the kidneys do what they do. You might ask: do mechanistic accounts of explanation capture everything going on here? Do these explanations constitutively involve what the mechanists say they do? Or is there something more going on?

*Draft 1De1. Compiled March 2, 2015

Completeness hasn't received nearly as much attention comprehensiveness. Professionally speaking, it doesn't seem like a sexy question. Charitable readers recognize that all accounts are incomplete. Filling in the details is the sort of cleanup job usually left for seminar papers, or done *en passant* on the way to some more interesting conclusion.

That's a shame, though. For I think that standard mechanistic accounts are incomplete, and importantly so. I will argue that not all mechanistic explanation involves citing components of mechanisms in the *explanans*. Some mechanistic explanations involve citing *resources* that are *used by* mechanisms. Resources can, and should, be distinguished from the mechanistic components that use them. That's not a mere tweak to the story.

I say more. One of the more persistent worries about mechanisms—present in many of the more daring critiques above, and a source of grumbling in conference hallways—is why we ought to be *mechanists* about explanation, as opposed to adopting a more general causal or interventionist story. Even if neuroscientists *do* give lots of mechanistic explanations, it might be that this is just an especially compact way of giving causal information.

Resource explanations shed light on this question as well. They'll do so by a rather indirect route: having made the distinction, I'll then argue that that the mechanism-resource distinction is itself not exhaustive. Instead, they are the most obvious results of an explanatory strategy that depends on the interests of particular disciplines. The interests upon which that categorization depends, though, are pretty important ones. Arguably, they're the sort that are constitutive of a scientific discipline. Because of that, the choice of mechanistic explanations is not mere convenience, but constitutes a vital explanatory strategy in the sciences that employ it.

2 Mechanistic Explanation

2.1 Parts and Wholes

I will focus on the account of mechanistic explanation given by Carl Craver in his (2007). A mechanism, in Craver's formulation, is "a set of entities and activities organized such that they exhibit the phenomenon to be explained." (p5). Mechanistic explanations are fundamentally multi-level. They work by demonstrating that the behavior of a whole can be captured in terms of the organized activities of its parts (p139).

Suppose I want to know how a car engine works—how it takes in fuel and air and uses them to produce radial motion in the crankshaft. A mechanistic explanation will break the engine down into components, explain what each component does, and show how it is related to the others. Having done so, I can show how the organized activities of the components produce the activity of the whole. I can show how the fuel rail is connected to the injectors, and delivers fuel to them, and then the injectors spray the fuel into the combustion chamber, and on and on. Once I have exhibited the activities of all of the components, one can see how the engine as a whole works.

Craver shows that this style of explanation is also found in many sciences. To explain the action potential, we show the channels embedded in the cell membrane, we explain the conditions under which those channels open, and so on.

A crucial aspect of mechanistic explanation is breaking whole objects down into their *mechanistic components*. A good account of mechanism ought to give a criterion that distinguishes true mechanistic components and their activities from mere spatiotemporal parts. If a wombat crawls into my engine compartment and goes to sleep, that doesn't make it a mechanistic component of my engine—even if, note, it might have a variety of causal interactions with the engine.

Craver appeals to the relationship of *constitutive relevance* in order to distinguish the true mechanistic components. Constitutive relevant is in turn cashed out in terms of mutual manipulability. Some proper part x of S is constitutively relevant to Ss ψ-ing if there is an intervention on xs ϕ-ing that would change Ss ψ-ing, and if there is an intervention on Ss ψ-ing that would change xs ϕ-ing (p. 153). In other words, it must be that one is able to reliably manipulate the behaviour of a constitutively relevant component to manipulate the behavior of the whole, and vice-versa. The behavior of the fuel injectors can be manipulated by speeding up the engine, and the engine can be manipulated by changing the speed of the injectors; that is why the injectors count as constitutively relevant. By contrast, there's no (reliable) change I can make in the interloping wombat by manipulating the engine, and poking the wombat has an inconsistent effect on the engine.

2.2 Abstracting Away

Mechanistic components are a diverse lot. Fuel injectors, hearts, sodium channels, and syntactic modules appear to have, on the face of it, little in common—other than that they are all cited in mechanistic explanations. Abstracting away a bit, though, we can see that mechanistic components share several important features.

First, mechanistic components *persist* over relatively long timescales — for now, let's say over the timescale of the behavior of the whole that we wish to explain. The different engine components are present the entire time the engine is running. That is an important feature of the mechanistic explanation: we implicitly assume that when the fuel rail sends gas to the injectors, the injectors are always there and ready to receive it.

Conversely, the failure of a mechanistic component is often the explanation for the failure of the whole mechanism to perform. The Challenger exploded rather than went into space because an O-ring failed. Tetrodotoxin kills by blocking the activity of sodium channels. Mechanistic components must thus be available and functional throughout the timescale of the explanation. Failure to do so explains deviations from the norm.

Second, mechanistic components are *individually important*. In addition to the *availability* of component-types, the *identity* of a component over time is often crucial to the explanation. My car starts hard in the morning. Why? It is too old to take ethanol-augmented gasoline. The ethanol destroyed the check valve in the fuel rail. As it sits overnight it loses pressure and makes for a hard start. Note that for this explanation to work, it is important that the same entity was affected by the ethanol, leaks overnight, and therefore explains what happens the next day.[1]

Thus mechanistic components don't just persist, but can persist as the very same entity over a variety of activities. This is obscured when we give mechanistic explanations because we often care about explaining *types* of mechanisms rather than individual tokens. But for any individual token mechanism, it is usually assumed that the same components persist, and persist as the same components, over time.

Third, mechanistic components are typically *casually conservative*. That is, each component interacts with only a limited subset of the other components in a small number of ways. The fuel rail does one thing, the intake another, the injectors a third; further, the fuel rail interacts only with the fuel line and the injectors, the injectors with the rail and the chamber, and so on. Casual conservativeness allows for *modularity*, which is a feature of complex mechanisms (Simon, 1996; Bechtel and Richardson, 2010). Further, mechanistic decomposition is made possible by the causal conservativeness

[1] Historical footnote: it was actually the starter.

of the individual components. Most of the processes we care about are complicated. We explain a complicated process by looking at the components involved and the limited sets of relationships between them. Note that the components need not be *simpler* than the whole. Rube Goldberg machines show that it's possible to build up a simple process from complex components. What's important, rather, is that each component affects something *less* than the whole, and so we can partition out the effects of components more easily.

Fourth and finally, our mechanistic explanations are often *indifferent to the composition* of the mechanistic components. Many mechanistic explanations care only about the location and characteristic activities of the mechanistic components. Mechanistic explanations often involve *functional* decompositions (Cummins, 1980, 1983), and functional decomposition cares only about the activities of the components and their relation to one another.

Note that this is a weaker claim than saying that mechanistic components are multiply realizable, at least in the classic sense. Mechanistic components might well be constrained, and rather severely, by the environment in which they must work (Shapiro, 2005). I'm inclined to think that myself. The claim is rather that for many mechanistic explanations, we don't care about how the components are composed: as far as the explanation is concerned, the fuel rail is just something that manages to get fuel from the pump to the injectors. The details aren't that important, at least when I'm explaining hard starts.

2.3 The Hegemony Thesis

The features common to mechanistic components illuminate some of the advantages of mechanistic explanation. Mechanistic components are persisting, repeatable aspects of causal chains that make a unique contribution to the behavior of the whole. Philosophers may differ on why things with those features are especially important features of good explanations—but few doubt that the *do* make for good explanations.

I will return to this point further in section 5. First, I want to point out a presupposition of accounts like Craver's—one that, if not stated outright, is strongly implied by things mechanists say. Let's use "mechanism" to refer to the whole whose behavior we intend to explain, and "mechanistic components" to refer to the spatiotemporal parts of that mechanism with the features just identified. The mechanists' claim, as I understand it, is that all good explanation of *mechanisms* can be cashed out in terms of the activities of the *mechanistic components*. Mechanistic components might not be the only things that contribute causally to the explanation of mechanisms. But they are (in ordinary circumstances) the only things we really *care* about when giving explanations of mechanisms.

So for example, mere background conditions also contribute causally to mechanisms. Engines probably won't work in the absence of gravity. Nevertheless, good explanations of engines can get by without mentioning gravity. Call this thesis – that appeal to mechanistic components does the sole and primary explanatory job in mechanistic explanations — *the hegemony thesis*.

The hegemony thesis explains why we ought to look for *mechanistic* explanations, rather than any old causal ones. Mechanistic explanations cite mechanistic components, which have a variety of attractive features. The hegemony thesis thus explains why mechanisms are so important.

Or it would if it were true. Alas, there are good *mechanistic* explanations of mechanisms that don't primarily cite mechanistic components.[2] It is to one species of those that I now turn.

3 Resources & Explanation

3.1 Some Examples

My engine stops running. Why? Because the car ran out of gasoline. Without gasoline, the engine doesn't go. That seems, intuitively, like a mechanistic explanation—certainly, it's on a par with the explanation of my engine's failure where I cited a bad check valve. (Many troubleshooting procedures for engine failure begin by checking whether you ran out of gasoline, before moving on to components that may have failed). Furthermore, gasoline seems to pass the constitutive relevance condition: by changing how much gas there is in the car I can change whether the engine goes, and running the engine changes how much gas there is.

Yet gasoline is very much unlike the rest of the engine. Gasoline is trans-

[2] Or, if you want to reserve "mechanistic explanation" for things that do primarily cite mechanistic components, that there are good explanations of mechanisms that look a lot like mechanistic explanations except that they don't cite mechanistic components. I find the former way—e.g. using "mechanistic explanations" to refer to the broader type—slightly less clunky. But this is a fight over an explanatory practice, and emphatically not over the term itself.

formed by the action of the engine. The engine persists through those transformations. Gasoline thus gets *used* by engines, while engines do the using.

To give it a name, gasoline is a *resource*. There are many resources that show up in explanations, across a variety of fields. The grass in the paddocks and water in the streams are resources for the livestock. The catalytic enzyme argininosuccinate synthase (ASS) is the rate-limiting step in the synthesis of arginine. Gas, grass, and ASS are thus resources that get used by the mechanisms of cars, cows, and mitochondria, respectively. They're also necessary: you can't ride for free.

Sometimes resources are familiar and concrete, as in the above. Other resources are more abstract. City planners must consider tax bases, storm drain capacity, and their ability to provide vital services. Each are resources that a city needs to run efficiently, but each is spread out over time and space. Computational complexity affects the amount of memory and processing time that an algorithm needs to run, and is in turn affected by the number of processors available to implement the algorithm. The performance of an internet connection depends on resources like bandwidth and downstream cache. Cognitive scientists appeal to a variety of resources: working memory capacity, attention, and willpower. We explain characteristic patterns of psychological performance and deficit by showing how these resources can be blocked, competed for, and otherwise made unavailable.

3.2 Features of Resources

Examples are easily multiplied (and more will come). I have already presented a diverse bunch. What do they have in common? I will suggest a cluster of features, each of which complements those of mechanistic components. As with mechanistic components, these are not meant to be necessary and sufficient conditions — that is not the game, and I'll explain later why I think such a search would be fruitless. For now, we'll just take in some ways in which resources are similar to each other.

First, tokens of resources rarely persist across the timespan of explanations that feature them. Cows eat grass. Engines burn gasoline. These activities irreversibly transform token chunks of resource, and that transformation is necessary for the proper functioning of the mechanism.

More broadly, resources can be made *available* or *unavailable* for transformation. The availability of resources is often a key feature in explanations. Working memory has a certain capacity, and holding items in working memory diminishes that capacity. Streaming a movie makes it harder to download system updates, because each takes a certain amount of bandwidth.

Second, and along the same lines, token chunks of resources are not individually important. That is, it doesn't (typically) matter which bit of resource gets used, so much as there's enough of it to go around. The car factory doesn't care which bit of steel gets used: it needs some ingot or other, but the particular one doesn't matter. Similarly it does not matter (for most applications) whether you use the first sector on the hard drive or the ten thousandth.

Mechanistic parts are individually important precisely because the tokens do such different things. Resources, by contrast, admit of equipotent divisions. Every bit of gas is basically as good as every other. These divisions might be effectively continuous (as with water supplies) or they might come naturally chunked (as with working memory). Each mechanistic part of the computer, by contrast, does something individually important—you can't interchange the hard drive for the memory.

Third, many resources are *causally promiscuous*, and perhaps all are intrinsically so. That is, resources can interact with many different components of a complex system, and many different components have an interest in using that resource. All of the grass in the paddock is available to any of the cows that want it. Even relatively restricted resources—like, say, modality-specific buffers—are available to any process that might use that resource type.

Causal promiscuity is worth noting because many resource explanations involve competition for a limited amount of resource. Conversely, many complex mechanisms have portions devoted to controlling access to resources. Multitasking operating systems spend a great deal of time mediating conflicts over access to memory and processor time. The brain promotes blood flow to areas that are working harder than usual (Nair, 2005). Even when resources aren't actually promiscuous, then, it is often because of deliberate strategies taken by mechanisms to *control* access, rather than the intrinsic properties of the resources themselves.

Fourth, we typically *do* care about the composition of resources, and otherwise similar resources can't be substituted willy-nilly. Car engines need *gasoline* – it would be nice if something else worked, but nothing works quite as well. Food chains are vulnerable to collapse precisely because organisms need specific resources. Ion channels might work in a variety of ways, but it's important that they let in only one specific sort of ion. Unlike mechanisms, then, we often *do* care about the composition of a resource.

Further, even substitutions that maintain functional similarity in a broad sense can cause a variety of problems. Ethanol-augmented gasoline in your old car, or a substandard batch of steel in your factory, may not shut things down — but that substitution is what you'll have to cite when you want to appeal why things break, fail, and otherwise fall short of expectations.

The contrasting features of resources and mechanistic components are summarized in table 1

Mechanistic components	**Resources**
Persisting	Non-persisting
Individually Important	Equipotent
Casually Conservative	Causally Promiscuous
Realization-irrelevant	Realization-relevant

Table 1: Contrasting components and Resources

Each of these contrasts should be read as relative rather than absolute. Sometimes we care about what mechanistic components are composed of, and sometimes the composition of a resource is irrelevant. Sometimes parts of a mechanism are destroyed in the course of action. On the whole, however, there are systematic differences between mechanistic components and resources.

3.3 Resource Explanations

Resources appear to be the primary *explanans* in a variety of mechanistic explanations. There are explanations that account for the activity of the whole in more or less the same way that canonical mechanistic explanations work, but which appeal to resources rather than mechanistic components. Again, examples will make the point.

Most obviously, many explanations *must* cite resources if they're to be complete. You don't know how a car factory works unless you know about the materials it uses and the machines that do the using. An explanation of how a heart works is obviously incomplete without mentioning the blood that gets pumped around. In many such explanations, resources might not play the most *interesting* causal-explanatory role, but they certainly play an *ineliminable* one.

Further, many explanations give resources pride of place. Some of these invoke various ways in which resources can be depleted, blocked, competed for, or otherwise made unavailable. Availability of resources thus shapes the behavior of the whole. Sometimes this relationship is crude: no gas, no driving. But the relationship can be more complex. Forming long-term memories via long-term potentiation uses glucose, which is largely provided by astrocytic glycogen. Depleting glycogen experimentally leads to amnesia. The elderly often have memory problems, and also tend to have lower levels of glucose. Supplementing them with glucose improves their memory (Messier, 2004; Newman et al., 2011). Crucially, the relationship between glucose and LTP is not a simple on-off sort of relationship: the relationship is a specific one, in the sense specified by (Woodward, 2010).

Computational complexity theory similarly posits specific relationships between an algorithm's properties and the amount of time, space, and so forth it can use. Consider: many low-level assembly language algorithms are tuned to avoid accessing RAM, which can be orders of magnitudes slower than accessing in-CPU cache (Duntemann, 2011). The performance of instances of the algorithms can be complex, but largely explained by the differences in access time between two distinct types of memory resource. Along the same lines, some enzymes are the rate-limiting step in reactions, and so the norm of reaction within the mechanism is largely explained by reference to the bulk availability of that enzyme.

Appeal to resources can also be evidentially important, because they serve to invalidate (or at least cast suspicion upon), certain types of mechanistic explanation. Tim Shallice's relies on considerations of resources in his argument that single dissociations offer only weak evidence for the distinctness of two cognitive processes (1988). Suppose I have a lesioned patient who can (say) spell regularly spelled English words and regular nonwords, but not irregular words. I might conclude that there are two routes to spelling: a preserved phonetic one that involves spelling-to-sound correspondences and a damaged lexical one that retrieves stored information about spelling. But as Shallice points out, there is another possible explanation: spelling irregular words might just be *harder*. That is, spelling might have only one route, but that route might require a general resource that is stressed more by harder tasks than by easier one – intelligence or working memory or attention, say. Partial damage to this resource will only affect more difficult tasks. Ruling out these *resource artifacts* is the primary reason why double dissociations are so prized in neuropsychology.

Note too that this appeal to resources and their properties are being done within the context of straightforwardly mechanistic box-and-arrow sorts of cognitive psychology. There's no need to go further afield to networks or dynamic systems to press the point. Even on its home turf, mechanistic explanation appears to leave something out.

Further, many of the techniques for distinguishing and exploring resources involve the same kinds of scientific methods that picking out mechanistic parts do. Just as it is not always obvious which spatiotemporal parts count as mechanistic parts, it is also not obvious which resources a mechanism actually uses. The very same techniques for exploring mutually manipulability that Craver details in his (2007) can be applied to resources. So for example, to determine whether something is a resource you can manipulate it, block access to it, change the whole to measure the amount of the resource, and so on. So it seems that scientists do very similar things to investigate both mechanistic parts and resources. The two thus deserve explanatory parity.

Finally, differences between resources can often explain differences between mechanisms. Sometimes these involve different resources as inputs. Different steelmaking processes, for example, vary importantly in what resources they use (forced air in the Bessemer process vs pure oxygen in contemporary BOS methods, say). Differences in resources often require different mechanisms: diesel engines don't have spark plugs and have heavier combustion chambers, for example, because diesel must be detonated via high compression rather than merely ignited. Explaining why the two types of mechanism differ thus cites the different resources those mechanisms are meant to use, even if they are functionally similar at the level of the whole.

At a more abstract level, resources themselves can vary in a variety of ways. As an incomplete taxonomy: resources can be limited or unlimited, they can admit of discrete or continuous divisions, they can be permanently transformed or merely blocked, usable in parallel or in serial, and so on. These distinctions make a difference to the systems that use them. Computers with random-access memory have different properties to older serial-access kinds. Omnivores have a different ecological profile than carnivores. Breakbulk shipping has a different timescale and efficiency curve to container shipping. The differences between these complex mechanisms is explained by the differences in the resources they use. Resource explanations are thus important when we *compare* mechanisms as well as when we try to account for the behavior of single mechanisms.

In sum, resources constitute their own class of *explanans*, distinct from mechanistic components, and sometimes forming the primary or even the sole explanatory variable for the behavior of mechanisms. The hegemony thesis is false: explaining mechanisms can involve more than discussing mechanistic components.

4 Interlude: Objections and Replies

4.1 Why Not Just be Broader?

The hegemony thesis states a strong restriction on the resources available to mechanistic explanation. One might object that it is unreasonably strong. Mechanists, the objection goes, have always included resources in their explanatory arsenal.[3] After all, Craver's mutual manipulability condition picks up resources too. William Bechtel has been especially clear about this, writing

> ...a part consists of components and operations. By *parts* I designate the structural components of a mechanism whereas by *operations* I refer to processes or changes involving the parts. I use the term *operation* to emphasize that in each operation not only is something performing an operation but something is operated on. In chemistry, investigators could identify the chemical substances (substrates) that are changed in a given reaction, and use that in characterizing the reaction. In the cognitive domain, investigators speak of information processing, suggesting that information is what is operated on... (Bechtel, 2008, 14)

Bechtel is clear: there are simply *parts* of a mechanism, and some of them are operated upon while others do the operating. Resources fall into the former category. Mechanistic explanation was never incomplete.

My response has several aspects. As far as sociology of philosophy of science goes, I submit that the attention of the mechanists has overwhelmingly been focused on mechanistic components. Even if some definitions managed to include resources, discussions of mechanistic explanations themselves almost always focus on the components doing the manipulating.

[3] Thanks to Elizabeth Schier for pressing me on this objection.

Second, as I've shown, resources have very different properties to mechanistic components. Thus, even if they are both (broadly construed) parts of mechanisms, there are good scientific reasons for treating them as distinct kinds of things.

Third, resources are not always spatiotemporal parts of mechanisms. Many resources — the grass in the paddocks, the water in the river, the bandwidth of the wifi network — are *external* to the mechanisms they help explain. They are not located within the mechanism, and it is unreasonable to extend the mechanism to include them. The boundary between mechanisms and the resources they use can be contentious—drawing that boundary in the case of cognition is at the center of the extended mind debate (Rupert, 2009). This is a further reason, then, to distinguish mechanisms and resources.

Finally, there is a good *philosophical* reason to distinguish mechanistic components. In section 1, I noted that a persistent problem for mechanists was explaining why *mechanisms* — as opposed to mere *causes* — were explanatorily important. The picture I sketched of mechanistic components gives one answer: mechanistic components tend to persist through a variety of different causal processes. Mechanistic components provide fixed points in the causal flux that can be related together to give an overall picture of the whole. Further, the same components can appear in a variety of different explanations, unifying them under a common ontological framework.

Resources, since they don't reliably persist, can't play the same justificatory role. On the other hand, they can play a different, distinct, and equally valuable role in explaining why mechanisms are the way they are. The realization-dependence of resources, for example, explains and thus anchors the form of mechanistic components. There are many ways one might build a check valve, but if it's a check valve for *high pressure gasoline* the available options are greatly reduced. This sort of information potentially gives depth and unity to otherwise disconnected causal explanations. Resources also vary much more frequently and variously than mechanistic components – in their presence, in their levels, in their composition, and so on. These variations are actual difference-makers across a wide variety of mechanisms, and so worthy of our attention.

In sum even if mechanists haven't distinguished components and resources, they have good reason to do so.

4.2 But aren't mechanistic components more basic?

A second objection notes the apparent *passivity* of resources. Mechanistic components sometimes work on their own. Clocks, those paradigms of mechanistic virtue, can be explained by reference only to their mechanistic components. Resources need mechanisms, on the other hand. Since resources are used, they need something to do the using. Mechanistic components are thus the more basic explanatory notion.

This objection must be phrased with care. For the issue is not whether resources are ever the *only* things appealed to in an explanation—for the sake of argument, let us grant that they aren't. Resources can still be the *most* important aspect of the explanation by being—for example—the only things that vary, and so the only actual difference-maker in a population of otherwise similar mechanisms.

The point could be read as a meta-philosophical one rather than one about particular explanations. Mechanistic components must be *conceptually* or *logically* or *explanatorily* prior to resources. We only care about resources, this line goes, because we care about the mechanistic components that use them.

This version also seems implausible to me. While *some* mechanistic components can be understood and appealed to without mentioning resources, many of them can't be. An explanation of how an engine worked that didn't mention gasoline would be hopelessly incomplete. Many mechanistic components can only be understood by understanding the sorts of resources they're meant to use: a fuel injector is something that squirts gasoline, and if you miss that you've missed the point. Even if not all mechanistic explanations involve resources, this wouldn't entail that resources aren't a vital and equal feature of the explanations in which they *do* appear.

A final way to run the objection appeals to different levels of explanation. So far, I've spoken of things as if they are either mechanistic components or else resources, and that this categorization can be made in some objective, atemporal sense. That is not so. The hierarchical decomposition strategy of mechanistic explanation often means that what is a mechanistic component in one explanatory context can be a mechanisms in another context. The capacities of the fuel rail can themselves be explained by treating it as a mechanism in its own right, and looking to its on mechanistic components.

Resources can also be treated as mechanisms. RAM is a resource for the programmer, but a collection of mechanistic components for the chip designer.

Computational neuroscience seeks to undercover the mechanisms that underly attention. Many purported resources, then, become mechanisms upon closer look. Perhaps that shows that mechanistic components are the more basic notion?

Alas, hierarchy cuts both ways. Mechanistic components can also be resources in some contexts. Kidneys are mechanistic components for the anatomist, a resource for the butcher. Jeeps are complex mechanisms to the motor pool, matériel to the general. Nor is there a general principle that the lower level is always the mechanistic one. RNA is a mechanism when you care about transcription, a cellular resource when you care about viral load. Both of those are explanations situated on roughly the same cellular-biological level. Water towers are (simple) mechanisms; part of the explanation of how they work treats the same object as including a very large spatial resource. More generally, the fact that some explanations of mechanisms appeal to resources (established in section 3.3) shows that sometimes resources can be lower-level explanatory aspects of a mechanistic component.

Mechanistic explanation, then, does not privilege mechanistic components as ontologically more basic or otherwise special. Mechanistic components and resources are equal partners in explanatory projects.

4.3 Aren't resources background conditions?

A final objection (and possibly the most pressing) seeks to assimilate resources to *background conditions*. On this line, resources do figure in mechanistic explanations. However, they do so only in a secondary or derivative sense, by providing the necessary background canvas upon which the real explanation can be sketched. Of course (the line goes) engines need gasoline to run, and wouldn't without. But they probably also wouldn't run in a vacuum, or in zero gravity, or in the presence of serious amounts of ionizing radiation. All of these specify background conditions that have to be in place before mechanistic explanations can get off the ground. But given that they are in place they – well, fade into the background. At best, they are appealed to to explain why mechanisms *fail* to work, not why they succeed. That's always a derivative project, though: failure is intelligible only against a background of success.

I suspect that something like this thought has occurred to many, and explains why resources have received comparatively less attention. I think the objection is misguided, thought. For starters, there is something more than a little odd about saying that you understand (say) how a factory works without understanding how it transforms the raw materials that come in. After all, that is the whole *point* of the factory. Car factories aren't complicated clockworks existing for their own sake, merely relying on a supply of steel to keep things moving. Every mechanistic component is geared towards transforming steel into cars, and you're surely missing something if you treat the presence of steel as a *mere* background condition. (Indeed, it seems to me more plausible to say that the presence of a factory is the background condition necessary for your factory to keep turning boring steel into shiny cars).

More technically, the standard marks of 'mere' background conditions don't seem to apply to many resources. First, resources can't be just bracketed off in mechanistic explanations. Some are, in Kenneth Waters' term, *actual* difference-makers (2007). That is, the level of some resource actually varies in populations we care about, and that actual variation is part of the explanation of actual differences between mechanisms in that population.

To return to a few cases: the same computational mechanism might vary in efficiency depending on how much memory it has available. The mechanisms underlying obesity are sensitive to differing proportions of carbohydrates and protein in food. Breathing rate and depth are determined by the level of carbon dioxide in the blood. Urea synthesis rates depend on relative availability of various enzymes. And so on and on. Note that all of these are explanations about facts of *normal* functioning. It's true that, in each case, the lack of the resource would cause the mechanism to fail. But that can hardly be a condition for explanatory irrelevance: pull a spark plug out of your engine and you'll get equally dramatic failure.

That leads directly to the second point. In a well-known piece, James Woodward notes that different causes might have more or less *specific* relationships to their effects (Woodward, 2010). Roughly speaking unspecific causes have few possible states that lead to few possible states of the outcome (at the limit case, just two states with a 'switch-like' relationship to the effect), while specific causes have many states that stand in a more or less one-to-one relationship to many distinct effect states. Woodward notes that when X is a relatively non-specific cause of Y, then "we are more likely to regard X as a mere enabling (or background) condition for Y" (Woodward, 2010, 317). Craver suggests a slightly different criterion for background conditions. A background condition, he notes, typically won't pass the *mutual* manipulability test. I might be able to affect my engine's functioning by changing the local gravitational field, but I can't change gravity by running the engine Craver (2007).

Both of these criteria seem to me like good rules of thumb for picking out mere background conditions. Many resources don't satisfy either of them. To recall one example: astrocytic glycogen is a crucial energy source for long-term memory formation. Forming memories depletes astrocytic glycogen, and depriving astrocytes of glycogen causes amnesia. So astrocytic glycogen satisfies Craver's mutual manipulability criterion. Further, there is a sensitive relationship between glycogen and performance on memory tasks Newman et al. (2011): a little depletion causes a little deficit, more causes more, and so on. So astrocytic glycogen passes Woodward's test as well.

Of course, not all resources will pass both tests (though I conjecture that most will pass at least one). There is a further, trickier question about how to deal with resources that do not coincide the spatiotemporal boundaries of the mechanisms (as with, for example, the ambient air), though I suspect those should also count as resources at least in some cases.

Despite these complications it's obvious that there's a deep *explanatory* difference between resources and mere background conditions. Indeed, the point can be turned around. I suspect that some of the difficulty in coming up with an adequate characterization of background conditions (at least for mechanistic explanations) might stem from the fact that mere background conditions are not reliably distinguished from resources. All the more reason, then, for philosophers to distinguish the two.

5 Mechanistic Explanation Reconsidered

To conclude, I'd like to return to the question that I raised at in section 1. Given that scientists in many disciplines apparently *do* care about mechanistic explanation, why is that the case? That is, why care about *mechanisms*, rather than just about *causes*?

I gestured at one possible answer in section 2.2. Mechanistic components tend to persist over time, and so explanations involving them have a certain kind of depth or unificatory power that would be lacking in explanations that cite 'mere' causes (whatever those would be). Still, one might find this less than completely satisfying. After all, *causal* explanations are also enhanced by describing stable features of the causal network[4]

[4] Just why that's an improvement is an open question. Most mechanists appeal to higher-order properties. i'm a fan of theories on which the benefits are pragmatic rather than metaphysical in nature; see my Klein (2014b) for some arguments.

We can say more. If it was *merely* a matter of persistence over time (say), resources wouldn't have the explanatory status that I've argued they actually have. Rather, I think the mechanist project is actually broader still. The goal of mechanistic explanation is to *taxonomize* causes (implicitly or explicitly), clustering them into groups with similar features. I've mentioned three such clusters so far: mechanistic components, resources, and background conditions.

Mechanistic components and resources, on the other hand, are distinguished by a variety of more fine-grained features. Again, these features are *explanation-relative*: a resource in cognitive science might be treated a mechanism in neuroscience. The carving is not arbitrary, however, and we can identify features of explanations that determine where an object falls in the taxonomy.

One obvious candidate for a taxonomic principle is the timescale of explanations. Processes that are effectively stable at one timescale can fluctuate at larger ones; conversely, fluctuating processes can exhibit longer-term stability at a coarser grain. Mental processes occur at a variety of timescales, from milliseconds to years (Newell, 1990, 80-81). One goal for mechanistic explanations is mechanisms is to find things that persist for some useful length of time, but "useful length of time" is obviously going to depend in part on the explanatory project.

Another candidate is the specific functional relationships in which an entity stands. RNA, for example, might count as a resource in explanations that treat it as an input to protein synthesis, but a mechanistic component when we treat (say) DNA as the input and RNA as a component in the mechanism.[5]

Put this way, there's nothing special about mechanistic components *per se*. Mechanistic components are explanatorily important precisely because they share features that make for good explanations. Resources similarly share distinct, but important, features, and these also make them good candidates for explanation. Their ongoing change and flux, for example, makes them excellent candidates for the actual difference-makers in a system. Further, even when they're stable, resources have a different modal profile than mechanistic components, and so support different kinds of explanations.

Perhaps the most abstract distinction, however, is that between the *agent* and the *patient* of action. Mechanistic components are agents: they *do* things. They might be *caused* to do things, but they are primarily active.

[5]Thanks to Chris Lean for this suggestion.

Resources, by contrast, are largely patients. They are acted upon. We speak about resources very naturally in the passive voice ("Gasoline was burned.") Their passivity is why we need mechanisms to control their availability and influence (Money doesn't do anything without people to move it around, banks to keep it safe, and so on). "Agent" and "patient" are, again, relative to the kinds of activity in question, which in turn leads to the explanation-relativity of the categories themselves.

Thinking of it this way makes clear that the categories of mechanistic component and resource might themselves be neither necessary nor exclusive. For disciplines without a clear agent-patient distinction, the two carvings might not get much purchase. Simple physical explanations are an obvious example: in the world of atoms and void, things simply *happen*, without actor or acted upon. This is why the mechanistic style of explanation gets little purchase in basic physics.

This is not something special about *physics*, though. Even more complex physical explanations might reproduce the agent-patient structure. Consider explanations of supernovas, for example: the star *uses* its fuel, and once it is forced to rely on inferior resources for stellar combustion, it goes nova. Having reconstructed the agent-patient structure, it also becomes natural to talk about the mechanisms of stellar evolution.

Conversely, there are systems that resist an easy agent-patient dichotomy because they appear to have features of both at once. The most obvious of these are what Bechtel calls *active* mechanisms (2008, Ch8). These include the mechanisms involved in autocatalytic loops, homeostatic processes, and other feedback cycles. Key to mechanisms with feedback is the possibility of a mechanism altering its own operation — that is, acting at the same time as agent and patient. Active mechanisms are, as Bechtel argues, key to understanding many living processes; they're also arguably more important for understanding many neural processes (Bechtel, 2008; Klein, 2014a).

Where does this leave us? I think we might conclude with a complex but realistic pluralism. Cataloging the things that go into mechanistic explanations — and, indeed, mechanisms, and cataloging things *as* mechanisms in the first place — is relative both to explanatory projects and how the world is structured. If that's true then we can see why mechanistic explanation is important but not fundamental. It is a *strategy* that some fields will employ more than others in the course of coming up with good causal explanations.

But that is not to downplay the importance of the mechanistic project. The details of mechanistic explanations will be heterogenous, and different projects may even come up with different ways of carving the world. Most obviously, disciplines which frequently encounter active mechanisms will have a more complicated time distinguishing pure mechanistic components, mere resources, and things that have features of both. Nevertheless, we have an answer to why we prefer mechanistic explanations when they are available, rather than just a causal/interventionist story. They pick out categories of *explanans* that share certain broad features that make for good explanations. I've suggested that there are more distinct categories than typically supposed, and that these deserve respect as well. Resources are an additional category that deserves special respect.[6]

References

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis, New York.

Bechtel, W. and Richardson, R. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT Press, Cambridge, mit press edition. edition.

Craver, C. (2007). *Explaining the brain*. Oxford University Press, USA.

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

Cummins, R. (1980). Functional analysis. In Block, N., editor, *Readings in the Philosophy of Psychology*, volume 1, pages 185–191. Harvard University Press, Cambridge, MA.

Cummins, R. (1983). *The nature of psychological explanation*. MIT Press, Cambridge.

Duntemann, J. (2011). *Assembly language step-by-step: Programming with Linux*. John Wiley & Sons.

Kaplan, D. M. and Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78(4):601–627.

Klein, C. (2012). Review of *Explaining the Brain*, by Carl F. Craver. *Mind*, 121(481):165–169.

[6] Thanks to Liz Irvine, Chris Lean, Elizabeth Schier, and participants at the 2014 Cognitive Science Kioloa workshop for helpful comments on an earlier draft.

Klein, C. (2014a). The brain at rest: What it is doing and why that matters. *Philosophy of Science*, 81:974–985.

Klein, C. (2014b). Psychological explanation, ontological commitment, and the semantic view of theories. In Sprevak, M. and Kallestrup, J., editors, *New Waves in Philosophy of Mind*, pages 208–225. Palgrave Macmillan, New York.

Levy, A. and Bechtel, W. (2013). Abstraction and the organization of mechanisms. *Philosophy of science*, 80(2):241–261.

Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67:1–23.

Messier, C. (2004). Glucose improvement of memory: a review. *European journal of pharmacology*, 490(1):33–57.

Nair, D. G. (2005). About being BOLD. *Brain Research Reviews*, 50:229–243.

Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge.

Newman, L. A., Korol, D. L., and Gold, P. E. (2011). Lactate produced by glycogenolysis in astrocytes regulates memory processing. *PloS One*, 6(12):1–11.

Piccinini, G. and Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3):283–311.

Rupert, R. (2009). *Cognitive Systems and the Extended Mind*. Oxford Univ Prress, New York.

Shallice, T. (1988). *From Neuropsychology to Mental Structure*. Cambridge University Press, Cambridge.

Shapiro, L. (2005). *The mind incarnate*. MIT Press, Cambridge.

Simon, H. A. (1996). *The Sciences of the Artificial*. MIT Press, Cambridge, 3rd edition.

Stepp, N., Chemero, A., and Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, 3(2):425–437.

Waters, C. K. (2007). Causes that make a difference. *The Journal of Philosophy*, 104(11):551–579.

Woodward, J. (2010). Causation in biology: Stability, specificity, and the choice of levels of explanation. *Biology & Philosophy*, 25(3):287–318.