DRAFT VERSION, comments welcome.

Explaining Mechanisms[1]

Peter Machamer
History and Philosophy of Science
University of Pittsburgh
pkmach@pitt.edu

1. Causality and Explanation

Not all explanations are causal. Often children and adults explain phenomena by categorizing, as for example when one categorizes a species of being of a genus. So a child may ask why does the chicken fly, and be quite informatively happy when told, "It's a bird." No doubt there are many other explanatory forms. Peter Achinstein called the categorization kind of explanation above a 'classification explanation'. He also discussed other kinds of explanations including special-case-of-a-law, identity, and derivation explanations. (Achinstein, 1983, Chapters 7 and 8). I will not talk about these. However, most of his time was spent on causal and functional causal explanations. I take functional explanations to be a type of causal explanation, and both of these I take to be best exemplified by explanations that describe mechanisms for how something comes about or is produced. I take this to be the most common type of explanations used in science.

In order to discuss causal explanations, in particular those causal explanations that are provided by mechanisms, we need to do a little ground clearing about causality. Ned Hall (2004) distinguishes two concepts of causation: dependence and production. So maybe not all causal explanations are productive. A version of the productive theory was the classical theory, where the major maxim until the 18th Century was: the effect is

[11] This paper owes much to many discussions with and helpful comments from Jim Bogen. Jim Woodward and Michael Strevens also helped me clarify points, when a version was presented in symposium with them at the University of Calgary. The Fellows visiting the Center for Philosophy of Science, Spring 2007, (Fritz Allhoff, Johannes Persson, Demetris Portides and Brad Wilson) and the Center's Director John Norton all helped shape this paper when they discussed it with me at one of their weekly meetings. Megan Delehanty (now a professor at Calgary) originally gave me a copy of the term paper from which I quote, back in 2001 when she wrote it for a seminar. She most graciously agreed to let me reproduce her diagram and use her example. I also learned much from Sandra Mitchell by listening to her lecture and then reading a draft of her book.

contained in the cause. Containment was their way of talking about the productive connection between cause and effect, often in terms of an active principle connected to the cause which brought about the effect. We don't talk like that now. Yet, maybe we should talk about causally productive activities, especially if dependence is not really a way of indicating causal continuity but only a way we talk about causes.

David Hume asked what is the necessary connection between cause and effect? He says that it's a fact that we cannot see the necessary connection. (Strangely, Carl Craver and William Bechtel (2007) take Hume to be making the claim that the cause and the effect must be logically independent. If this just means there is no necessary connection, then it adds nothing to the no-necessity claim. If it means something more, like there are no 'logical' relations among the entities that stand in a cause effect relation, either I do not understand it because I don't know what logical relations are between entities or it's false because we can always find numerous relations among entities satisfying most any criteria.) But Hume was right, we cannot see the necessity, though I distrust facts about what cannot be seen. So he's right, though not in the way he thinks. We cannot see the necessary connection, because there are no necessary connections in nature. So it might seem we should be dependence theorists. But this would be to argue that the way we sometimes think and talk is really the way things are, to confuse the ontic with the epistemological. Jim Woodward's (2005) idea that causality makes a difference to what happens can be read either epistemically or ontically. He uses it epistemically when he talks about intervention as the way to find out about causes, and tries for an ontic claim when he invokes counterfactuals. I would cash out "make a difference' is terms of causal productivity.

In terms of ontology, I believe everyone today agrees that nature is contingent (though, for example, Hobbes and Spinoza did not). The world didn't have to be the way it is. Yet it does not follow that all is accidental or merely correlated. "Accidental" does not mean the same as "contingent". Some contingent connections are causal. Others are not. Once one gets rid of this idea that in the world there exists some sort of necessity for the causal, the major metaphysical impetus for thinking about causality in terms of counterfactuals disappears. Hypothetical conditionals, subjunctives, and future conditionals will do all the work we need for planning, intervening, discovering, and testing. And we can and do make contingent means-ends claims about hypothetical necessities. But this is not, nor do we need to have, a metaphysical theory about necessity based on counterfactuals and/or possible worlds. We do make general metaphysical claims when we speak of entities engaging in productive activities. Yet such a way of talking does not constitute a theory. Maybe it could, but who would be interested?

In the rational world of persons (the space of reasons), there may be good reasons to speak some times of 'necessary' connections such as truth preserving inferences in logic. Maybe these are necessary, where "necessity" means something like "on pain of contradiction".   (But is contradiction really such a pain? Who wants a foolish consistency any way? John Wisdom once said that if all you can do with a philosophical argument is twit your opponent with inconsistency, this is not very interesting.) But let's ignore the complications of reason for the moment.

Connections in the natural world come in many varieties. We have plethora of verbs to describe them. Mere correlation, however, is not a connection. It presupposes a strong independence between the 'things' correlated. It is only an association between two or more 'things', a pairing. It may be accidental, meaning that we understand no reason for the connection, or that there is no causal relation among the 'things'. If one is a strict associationist in philosophy or psychology, then this is all we can have. But even Hume saw that such pairings don't lead us to act as if this all we have. Maybe there is good reason why we don't.

Another constraint on connections is that we would like them to be relatively regular, or at least be possibly repeatable in something like the same ways. It may be that a mechanism or causal connection, in fact, works only once. Think of Rube Goldberg machine that self-destructs. But if it was a causal connection, and not accidental or a mere correlation, we would be willing to bet it could happen again. Sandra Mitchell (2008) has described this criterion of connection in terms of *stability.* Cf. Woodward and Craver on invariance, perhaps Wimsatt robustness.

Causal connections may not be necessary, but they do exhibit a continuity that provides the basis for why we find them to be intelligible. By discovering the kinds of connections, we can give the reasons why they are ordered in connected ways. In science, acceptable causal relations are those that our scientific investigations reveal to us as how the world works. We have to discover them. However, what are taken as acceptable connections, what we think we have learned, changes over time. Acceptable bases, what are called “bottom out” fundamentals (MDC 2000), are historically domain or discipline dependent. Yet again, let me caution that “bottom out” while suggesting down, need not be a reductive strategy. The bottom out fundamentals at a time need not be ‘lower level’ mechanisms.

Many years back, Norwood Russell Hanson (1958) wrote: “The primary reason for referring to a cause of *x* is to explain *x*. There are as many causes of x as there are explanations of *x*. …We have had an explanation of *x* only when we set it into an interlocking pattern of concepts about other things, *y* and *z*. …what we refer to as ‘causes’ are theory laden from beginning to end. These are not simple tangible links in the chain of sense experience, but rather details in an intricate pattern of concepts.” (Hanson 1958, p. 54). His example is seeing a stab wound, and knowing what kind of thing must be its cause and the general background conditions in which describing a phenomenon as a “stab wound” makes sense. Later, he adds: “The background information, the ‘set’ that makes an explanation stand out, derives as much from what is obvious in the situation as from discursive knowledge gained through training” (p. 62). Hanson is a causal pluralist (Cf. Mitchell 2009 & 2003). Earlier, Stephen Toulmin (1953) had made a similar Wittgensteinian point, when he described an explanation as being like a map. More recently, Michael Strevens (2007) acknowledged some of these connections when he wrote: “Because the mechanical information is subject to constraint from below, the search for a mechanism is guided in part by wider beliefs about the workings of the appropriate basic level.”

Background information, wider beliefs, and such ways of talking point to the fact that we explain phenomena using terms that connect what is to be explained with other things that we take to be more basic, fundamental, and which we think we know. Note that “more basic” or “fundamental” need not be taken in any ‘reductive’ sense. It maybe that what is fundamental to being a person is to make inferences in such a way that one is responsible for their conclusions or the actions that follow from them. We would then have to explain this inference making ability and responsibility ascription, at least in part, by tying these activities to social parameters and conditions.

This tying together of what we already know or take as knowledge sounds much like the old familiarity thesis. We explain phenomena in terms of what is familiar. Or, phenomena are explained when they are 'reduced' to the familiar. Carl Hempel and Paul Oppenheim (1948) argued against one version of this thesis, viz. that the explanation had to produce some feeling a familiarity in the person who understood the explanation. (*Aspects*, p. 257). But the Wittgensteinian and Salmonian move to inclusion in an (ordered) set of concepts and how these refer to acceptable causal relations among the things of the world does not depend on psychological feeling, so this form of the objection does not touch our ontic version of familiarity as background knowledge.

In explanations we do explain by showing how the phenomenon came about in terms of what we already accept about how things work. (This sounds almost like Mackie's 'laws of working' (1980), but seldom are they what we describe as laws). So in this sense scientists do explain things in terms of what is familiar to them, by training (which depends on field or discipline). They learn causal exemplars. Schaffner (in correspondence) called them 'accepted causal prototypes'. There is of course, one notable, most important exception; when one discovers a novel explanation by discovering a new mechanism or mechanical process and changes what a discipline believes. In such a case the reformer introduces new, unfamiliar causes to provide the explanation. But even here, the novel explainers must conform to many other things we know about how things work, what causes what, etc. Presumably this is what Watson and Crick did when they described DNA as a double helix that unwound to produce RNA (and eventually proteins).

I have already shifted from causes to explanations. So let us continue. Here are three claims I wish to defend, given the restriction above to a limited kind of explanation (i.e. to causal, functional explanations):

1. Explanations should provide causes (or reasons).
2. Explanations should make phenomena intelligible.
3. Explanations should exhibit the continuity among the explaining parts.

I shall elaborate each of these claims as follows:

1'. Explanations provide causes by providing mechanisms. Knowledge of mechanisms provides us the with the reasons we use in explanations (and in the self referential case of inference, there is a mechanism by which reasoned inference works, albeit a social-cognitive mechanism.)

2'. Intelligibility (or understanding, if you prefer) is brought about when one sees how the phenomenon is produced according to a familiar, acceptable set of more basic beliefs, or how a mechanism, and entities and activities that comprise them, produces the phenomenon of interest in accord with some more basic beliefs. (Note these basic beliefs need not be at a 'lower' level. Sometimes we explain why a phenomenon is produced by appeal to larger systems within which that mechanism functions.) Most if not all explanations are multi-level. For many, if not most cases, there are multiple explanations that are not alternatives to each other. So, it is not the case that there must be only one acceptable explanation for any phenomenon.

3'. Intelligibility depends crucially on comprehending the continuity among the entities and activities that comprise the mechanism that is the explanation. This continuity is not the unifying of domains spoken of by Michael Friedman and Philip Kitcher. It is the continuity or sense that is inherent in the already accepted relations among entities and their activities that make for a coherent narrative. This why they may be used as part of a mechanism to explain a stage in the process about how a mechanism produces a phenomenon. I am also ignoring here a difficult point, discussed at length by Achinstein (1983), about *who* can make *what* intelligible *to whom*, i.e., the pragmatics of someone giving explanations *to* someone else. In fact, this is a topic for experimental philosophy.

Let us still briefly look at each of these claims in turn.

2. Mechanisms:

Over 40 years ago, Peter Alexander (1963), though he did not use the word "mechanism" wrote: "What is needed for an explanation of the observed movements [of a toy engine] is an account of something other than these movements together with the exhibiting of a certain kind of connection between this something other and those movements. The sort of connection required is roughly indicated by saying that what is needed is an account of whatever makes the movements in question occur or of something of which these movements are a consequence." (Alexander, 1963, p. 117). He goes on to say that we need to see how the explaining features are relevant to what is being explained Further, "it is also seeing that the arrangement of the parts is such that when the wheels revolve the head must move in the way it does, in some sense of must; it is seeing how the movement is a consequence of the arrangement of parts. " (Alexander, 1963, 118). Alexander is still hung up on the necessity, and so the emphasis is on the *must*. But it seems quite clear he is describing a mechanism. He also says "It [the explanation of the toy engine] would be acceptable to most people because it is in terms of pulls and pushes." (Alexander 1963, 116).

A short time later, Rom Harré (1970) "In order to give a scientific explanation every happening must be looked upon as due to the workings of some mechanism, which may be proceeding in isolation from its environment…or may be in various degrees dependent for stimulation in the circumnambient conditions…" (Harré 1970, p. 124) Earlier he (Harré 1961) had entitled a chapter of his book *Theories and Things*, "From Models to Mechanisms". So there is an interesting history of tying mechanism to explanation.

In Machamer, Darden and Craver (2000), we cited William Wimsatt, Kenneth Schaffner, William Bechtel and Robert Richardson, and Stuart Glennan as persons who had explored the concept of mechanism. Wesley Salmon and Phil Dowe also have mechanistic theories. More recently Michael Strevens is defending an account of mechanism explanations (Strevens 2008). MDC defined a mechanism as follows: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions." (MDC 2000). One problem with this definition is the inclusion of the word "regular". Some mechanisms are not regular, and indeed, there may be mechanisms that work only once. There also may be more than one mechanism for producing the same effect.

William Bechtel and Adele Abrahamsen (2005) give brief definition:

> 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. (2006, p.3)

I guess here "responsible" and "functioning" are the causal surrogates. Further, mechanisms are not just structures, they act and do things.

Recently Jim Bogen and I wrote:

> A mechanism is a collection of entities that act in certain causally specific ways. When a mechanism operates successfully, the entities in the collection engage in a temporally ordered sequence of activities by which the mechanism moves from an initial to a final state constituted by the production of one or more effects or end states. The entities in the system are ordered spatially, temporally, and functionally as required for the productive continuity of the activities that bring about the end state.

Explanations provide information that shows why or how an event or state of affairs came to be[2]. The operation of a mechanism is how or why a state of affairs came to be. The information an explanation provides by describing a mechanism is information for use in inferences about how the mechanism will continue, how it reaches it's end state, and how certain stages by virtue of their activities relate to earlier or later stages. In this way mechanisms provide information about what does, could or would happen if the mechanisms is not interfered with or undermined in some way. Such interference (or interventions) are often the best way to find out what are the parts of the mechanism and what is causing what to happen in it.)

Sometimes mechanisms are used to explain by showing how they function in 'larger' systems. Evolutionary explanations of specific mechanisms for the emotion of fear of snakes in terms of natural selection are one example (see Roach, 2001). So is explaining why a person's action, e.g. stopping at traffic lights, is seen as acceptable in terms of social (in the example case legal) norms.

---

[2] This is an inferential sense of information, which will be described in more detail below.

3. Intelligibility

Strevens (2007) touches on the intelligibility question when he speaks about "the constraint from below". He says any mechanism must be implemented at some appropriate basic level, and what are appropriate levels are field dependent, not metaphysically fundamental and not known apodictically (i.e., they can change.)

In MDC (2000) we wrote:

> Nested hierarchical descriptions of mechanisms typically *bottom out* in lowest level mechanisms. These are the components that are accepted as relatively fundamental or taken to be unproblematic for the purposes of a given scientist, research group, or field. Bottoming out is relative: Different types of entities and activities are where a given field stops when constructing mechanisms. The explanation comes to an end, and description of lower level mechanisms would be irrelevant to their interests. Also, scientific training is often concentrated at or around certain levels of mechanisms. Neurobiologists with different theoretical or experimental interests bottom out in different types of entities and activities. Some neurobiologists are primarily interested in behaviors of organisms, some are primarily interested in the activities of molecules composing nerves cells, and others devote their attention to phenomena in between. The fields of molecular biology and neurobiology… do not typically regress to the quantum level to talk about the activities of, e.g., chemical bonding. … But remember, what is considered the bottom out level may change.
>
> These bottom out activities in molecular biology and neurobiology can be categorized into four types:
>
> (i) geometrico-mechanical;
>
> (ii) electro-chemical;
>
> (iii) energetic;
>
> (iv) electro-magnetic."

The adequacy of the taxonomy of these 4 types of activities has not been explored. But they are the types, or at least some of them, that characterize mechanisms, and their workings and which are taken as being intelligible. Carla Fehr (2005) argues that "The explanatory power of a mechanism is dependent upon its epistemological context and the contexts of these mechanisms differ." She wants to 'unpack' epistemological context in terms of Helen Longino's local epistemologies. Certainly what is accepted is relative to a group, and maybe fields and disciplines are too large in many cases, especially fields that break into subfields or disciplines engage in controversies.

When a scientist in neurobiology talks about an action potential's arriving thereby raising the membrane voltage sufficiently for opening an calcium specific ion channel, this is taken to be an accepted mechanical description. It was not always so. Carl Craver describes how in 1969, Bertil Hille who was working out the mechanism for Hodgkin and Huxley on action potentials was chastised by Kenneth Cole for pushing the idea of channels too far. (Craver 2007, chapter 2).

This intelligibility is somehow transitive in explanatory contexts. The lower levels, described as stages in a mechanism, do not merely constrain what happens, they are the explanation of how what happens happens. That they must be able to be implemented in known or conjectured physical entities and activities serves as a constraint on providing an explanation by a mechanism. And often a scientist comes up with a possible mechanism, before finding one (that one or a different one) that is plausible or actual. It needs also to be noted that each of these explanations by mechanisms is multi-level and so not reductive in any sense. Further, it should be understood that mechanisms may be identified and explained by 'higher' levels and need not always be given in terms of decomposed, constitutive parts. An information processing mechanism is one such case because the information is for use by some system that contains that mechanism.

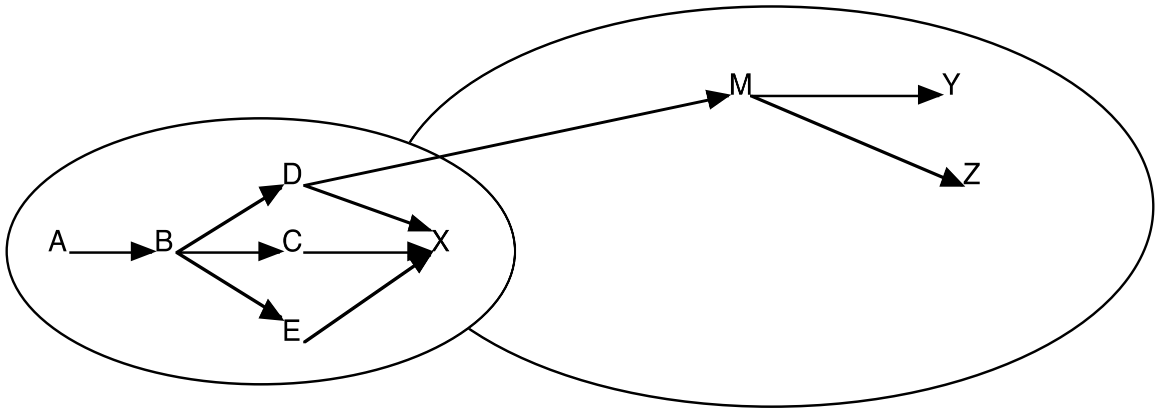
Yet there are other important kinds of constraint. I shall only explore one. Carl Craver wrote in recent draft of a paper, "The boundaries of what counts as a mechanism (what is inside and what is outside) are fixed by reference to the phenomenon to be explained" (Craver 2007, and he cites Bechtel and Richardson 1993 and Glennan 1996, and see Bechtel 2008.) But "fixed" is most often too strong. There are many ways to skin a cat; and, the same protein may be made in so many different ways that it often becomes impossible to form a generalization (or law) that there is a single 'gene' responsible, or that the mechanism used in a particular case is always *the* mechanism for making that protein. (Griffiths and Stotz, http://www.representinggenes.org/, and more generally for pluralism, Mitchell 2003). In most any case in neuroscience, the termination conditions allow for many alternative paths that would bring them into being. Carl Hempel (1965) recognized this when talking about functional explanations when he noted that they lack necessity because of alternative mechanisms for achieving the same end.) So for example, pleasure may be obtained in numerous ways, resulting from many different mechanisms. But looked at from the start up conditions forward, we are seeking a causal specificity that exhibits why this path was the one in this case that brought about the end state (Cf. Woodward 2007).

An example will help specify these claims, and further allow us to bring out some different points. A classic textbook on *The Central Nervous System* (Brodal 2004) explains part of the olfaction system as follows:

> Experiments with a large number of ordorants suggest that the *shape of the molecule*, rather than its chemical composition, determines how it smells (stereospecificity). The *stereochemcial theory* of smell proposes that the receptor sites on receptor cells have different shapes and that only molecules with a complementary shape fit into that receptor site. Binding of an ordorant to receptors in the membrane of the cilia evokes a *receptor potential*. This involves activation of G proteins and cyclic AMP. Increased intracellular cyclic AMP opens Na+-selective cation channels and thus depolarizes the cell, corresponding to the situation in photoreceptors. In contrast to photoreceptors, however, the olfactory receptors produce action potentials. The action potential arises in the initial segment of the axon and is transmitted to the olfactory bulb. (p. 234)

Here there is described a mechanism (*albeit* probable) that explains how an odorant (or some kind of molecule that produces a particular smell) binds to a receptor site and evokes an action potential which explains the end state, here the arrival of the smell signal in the olfactory bulb. The explanation proceeds by breaking down the olfactory system into composing entities (molecules, receptor sites, cyclic AMP, olfactory bulb, etc.), their properties (shape) and their activities (binding depolarization producing). (For typical decomposition strategies see William Bechtel and Robert Richardson 1993.) This part of the olfaction system is identified by describing how the mechanism works in order to send the smell signal into the olfactory bulb. Needless to say the explanation in the text does not stop here. It goes on to describe how the efferent fibers from olfactory bulb end at the medial aspect of the temporal lobe—partly in the cortex and partly in the amygdala. Also of interest is that mechanism as described, in its startup condition, as beginning with an odorant. Obviously there is more that might be said about how odors enter the nose, get dissolved in the mucous, and become separated into different molecules.

This demonstrates that what counts as a mechanism depends on the end state chosen. Where the explainer decides to stop. This is a partial function of the purposes or interests of the investigator. We might call such ending points, *perspectival ends*. In a typical case, the investigator identifies the system in terms of its end state, which is the end selected because it is relevant to her research. The system and its mechanism is simplified, idealized, and considered to be isolated or closed, and independent of other systems. But this perspectival end state is only one aspect of exploring mechanisms. More significant constraints come from what is discovered about the world. The anatomical and physiological studies provide knowledge about the world that constrain where the investigator looks, what she studies and tries to isolate and identify, what can be discovered, and, most importantly, what end states may be chosen to be investigated. These are objective constraints even given the scientists' choices. These objective constraints may be studied by trying to ascertain the causal specificity among the parts and activities being studied (Woodward 2007). I'll call these *natural ends*. These perspectival and natural ends contrast with a stronger more robust notion of teleology that is at work in some mechanisms, particularly those that are said to carry information. But more about information later.



adapted from Megan Delehanty, mss.

Here's a diagrammatic example (which I have taken from a term paper written in 2001 by Megan Delehanty). If the scientist is interested only whether or how X is produced, then that fact that X may be produced in any of three ways, or in combination, from D, C or E pathways does not tell us more than X is produced and that it came about in any one of the three ways. We do not need to know which of the pathways was active. However if we look farther and are interested now only in X, but also in its relationship to Z, then it matters that X was produced by the activity of the B-D pathway. Here what is identified as the relevant mechanism clearly depends upon what our interest is. Delehanty actually provides a real case that fits this schema, that of *ras* genes in *Dictyostelium*. She writes

> "There are five *ras* genes in this organism, expressed at different but overlapping times in development. *RasB*, *rasC*, and *rasS* are all expressed during the amoeboid phase as well as during aggregation and are redundant in regard to their function in aggregation. If we take aggregation as our end point, it is not essential to an explanation of why aggregation proceeded normally to know whether *rasB*, *rasC*, or *rasS* was active (whether, say one had been knocked out.) If we are interested in aggregation not only as a phenomenon which is independently interesting, but as one of set of activities requiring regulation of the cytoskeleton…we might care that *rasB* has effects on the cytoskeleton which are not shared by *rasC* or *rasS* and which do not affect aggregation but do cause defects in cytokinesis (Z)." (Delehanty, mss. 2001; quoted with permission.)

Mechanistic descriptions such as the olfactory system or *Dictyostelium* are a prelude to working out procedures for the discovery of mechanisms. (Lindley Darden and Carl Craver have worked on this problem, Craver 2001 & 2002; Darden 2007.) Notice though that the end conditions, evoking an action potential that travels to the olfactory bulb or aggregation and defects in cytokinesis, are taken to be unproblematic natural end points. In the smell case it is natural because we have long believed we know many of the parts and activities of the mechanism, and we are just trying to establish more clearly how it works. In aggregation we have come to know that this is an important part of what the organism does. Yet, there is no incompatibility in these cases between being natural end point and being socially or personally chosen. These ends are both perspectival and natural. The researcher identifies the perspectival end stage of the mechanism (or system), although it is a naturally occurring outcome of a production in nature. It a natural state treated as an end. (Marcel Weber (in correspondence) noted this dual aspect.) The researcher also has to make many decisions as to what to include in the mechanism that produces the end state, and these decisions too are constrained by what's there, what's happening, and what techniques she has available for investigation. One way to think about the need for such choices in naturalistic contexts is that there are no closed systems in nature, and so the researcher must put limits on them (close them) for explanatory and investigatory purposes. This echoes the rationale given for controlled experiments.

There are some limits usually provided by the physiological constraints of what we believe we know about the system under study—our background knowledge or beliefs. More philosophically, we generally do not countenance mechanism explanations for 'non natural,' Cambridge complex properties (e.g., the object or event formed my mixing a martini today at 4 p.m. and Jim's fear of his visit to dentist tomorrow.) We do not think we could ever find any productive connection among these activities. We have no useable criteria or background knowledge that places these events into a unified system. Our presumed knowledge at a time is used to specify what counts as a system as well as what is included in that system to be studied. If we wish to study some emotional mechanism, our knowledge (borne, as Tom Kuhn said, from our training) of what constitutes the emotion systems puts boundaries on where and what kinds of mechanisms, entities and activities we may look for. This was the point of incorporation into a system of concepts that Toulmin and Hanson were theorizing. Yet, at times we may decide to question that 'accepted' knowledge.

What count as proper phenomena for explanations by mechanisms depend, most usually, on the criteria of importance that are in place in the discipline at a time. This means there is always an evaluative component operative in selection of end states to be explained. Yet, there are physical systems with equilibrium states, where being in equilibrium or in homeostasis is the natural end, and we seek to discover the mechanism by which such equilibrium is achieved. Here we might be tempted to say that this is a natural end system, where the importance of the end is given by nature, and we just become interested in it. But if somehow we established there were no true equilibrium systems in nature, or that the single phenomenon we are investigating is really myriad of different systems operating in different ways, then it is we who would have made the mistake in identifying the phenomenon of our research and treating it as real, though describing such a state for comparison's sake could still prove useful. This is the problem of external or ecological validity.

4. Explanation and Continuity

The continuity I am concerned with lies in the intelligible connections that unify or connect the stages of mechanisms. Intelligible connections as well as acceptable entities, as we saw above, are assumed as acceptable, and so constitute the tool box, the explanatory text, that may be used in a domain (or field) at a time. Maybe it would be helpful to describe this in terms of the explanatory narratives that define a field? Sometimes intelligible connections (mechanisms, mechanism schema, or mechanism stages) in one domain are extended by analogy or model extension to another, new domain; and this may unite two domains to be explained into one, so this would be how we could attach the continuity I speak of with the unity of Michael Friedman and Philip Kitcher.

Causal connections or activities provide the continuity among entities (or stages) that are constitutive of the mechanism. This continuity, perhaps, is what people sought when they were looking for the Hume's "necessary connection". The continuity shows us why what happens brings about the goal state of interest. This in turn is the reason why the mechanism makes a difference; the difference being bringing about a specific end state. The continuity comes by means of the activities (represented by transitive verbs or participial expressions), which show how the entities involved act in their manner of production or by what action they connect to the subsequent stage.

The mechanism by which vinegar is made from wine is that a group of bacteria (called "Acetobacter") converts the ethyl alcohol portion of the wine to acid. This is the acetic, or acid fermentation, that produces vinegar. Bacteria converting alcohol involves two entities and a connecting activity. The end state is the vinegar. There are more detailed ways of specifying the mechanism, for example:

2 CH3CH2OH + 2 O2 --- > 2 CH3COOH + 2 H2O
Alcohol + Oxygen ----> Acetic Acid + Water

In these two schemata the arrows represent the activity of production or conversion.

The continuity in the mechanism, as a whole or among the parts, is represented in our descriptions of the mechanism by a kind of semantic connection that shows how nouns (usually) are connected by verbs, where the connection is causal. Presumably this is what Hanson (1958) was talking about when he drew attention to the semantics of 'puncture wound'. (In a couple of papers William Ruddick (1968) tried to elaborate this kind of 'semantic' connection.)

Some recent linguistic work, has noticed this continuity producing function of the ways verbs work. Verbs, from my point of view, can be taken as linguistic representations of activities. Some cognitive grammar people represent them as event schemata. So for example, W. L Chafe (1970) argues

> The nature of the verb determines what nouns will accompany it, what the relation of these nouns to it will be and how these nouns will be semantically specified. For example if the verb is specified as an action, as in *The men laughed*, such a verb dictates that the noun to be related is *agent* which might be further specified as animate. (http://www.ilc.cnr.it/EAGLES96/rep2/node10.html)

Or again, describing the changes in linguistics in the late 1980s and 1990s, the *Encyclopedia of Linguistics* speaks of the introduction of function words that divided sentence structure into functional domains. The first of which is "a lexical domain around the verb, which establishes semantic relations between the main sentence elements." (http://strazny.com/encyclopedia/sample-function-words.html; p. 2)

Transitive verbs connect subjects with predicate objects or events. When these subjects and objects refer respectively to prior and posterior stages in a mechanism, the verb describes the activity by which the one effects or produces the other, and so they are our way of expressing the intelligible connections between the stages. Verbs (activities) connect the referring nouns, and allow for a description of the produced stage that may serve as a prediction as to what will come from that stage (or backwards what occurred that allowed that stage). Choice of a particular verb, from the explanatory toolbox, in general constrains the acceptable descriptions of the next stage. In "The bacteria convert the alcohol", the 'convert' describes the production process by which 2 CH3CH2OH in the presence of oxygen (+ 2 O2) turns into 2 CH3COOH in its next stage. Bacterial conversion tells us what to expect from the open bottle of wine that we left out all night. We might say, in the spirit of contrastive explanation, that the activity(-ies) and the originating entity(-ies) select by constraining the range of possible outcomes. We would not expect, under normal circumstances, the formation of lactic acid formation, C3H6O3 or tartaric C4H6O6. This seems to coincide with Woodward's causal specificity as ruling out alternative outcomes.

This view of verb function contrasts an older view of D. Genter, (1978), who treats verbs as relational between nouns. She holds that since only nouns refer, verbs taken as relations must be abstract. So she believes nouns are processed easier and earlier than verbs. This is to treat verbs are logical relations, rather than as referring terms. Logical relations are the wrong way to look at verbs (and nouns). This verb function has been recognized and elaborated upon by Michael Tomasello ad Jean Mandler.

Intelligible connections, and their associated descriptions, are learned as specific kinds of causal connections, most often from particular prototypical cases or exemplars. This view of the centrality of activities fits G. E. M. Anscombe's theory of causation, where there is no deep metaphysical truth to be had about the nature of causality, but only specific cases and kinds that are learned and so recognized as causal.

5. Continuity and information

There is another kind of unity, common in biology and neuroscience in which the flow of information is used to describe what a system does in a way that shows how the stages are connected. “Information” is used in such explanations to provide an overarching connection among the stages, thus providing a principle of continuity. Information attribution provides an identity condition (same information) that unites the stages as having contents that causally function as stages working towards the same end or as being guided by their beginning stage.

Some cases of information are like the information conveyed by a cooking recipe in that they provide instructions that guide subsequent inferences or actions (See Stegman 2005, pp. 435-6). Another kind of information functions like evidence that directs us towards drawing a certain inference. This is a kind of information that Ruth Millikan (2003) speaks about. She also talks about teleosemantic information which directs processes that satisfy a goal set by evolution. Yet another common use of information comes from communication theory, and is often called Shannon Weaver information, which is a mathematical relation between a sender and a receiver defined as a probability relation among discrete elements of a message. I am not here interested in any of these.

Jim Bogen and I have been trying to characterize a different kind of information that we call ‘mechanistic information.’ This is what we believe is common in the biological and neuro-sciences. For example, an organism’s system triggers some bit of DNA to ‘make a protein’ that is somehow needed by the larger system. The structure of the DNA and the activity of unwinding produce a guide for the production of the next step. The structure of the DNA is a constraint on what RNA can be produced. Farther down stream, these activities constrain the amino acids that can be gathered to form the polypeptide. DNA producing polypeptides is not the result of human intention like following a recipe or drawing an inference. Yet it almost seems as though we are attributing a sort of intentionality, or specific kind of productive causal role, to the DNA by virtue of its structure. The DNA ‘code’ functions to direct the stages of the mechanism that puts together the specific sequence of amino acids. DNA makes the polypeptide as an end product, and this end is a goal state or privileged because the system needs it to go onto make the protein. The protein fulfills some need of the system. The organism needs to protein in order to, for example, maintain its metabolic functioning. In general the purposes information serves need not be, though sometimes it may be, a function of evolution by natural selection. They also may be end states that have been learned to be important (learned goals as Millikan says). But often the goal is only an end state that the system needs to be able to function and/or maintain itself, or a goal that we, theorists, have identified as interesting for us among the various ones that are produced by the system. These are all sufficient end states for a teleological system

But this teleology is not semantic in that the stages cannot be described as having semantic content in any interesting sense. They have by virtue of being a mechanism the same kind of continuity that all mechanisms have. But there is more. The stages are simply parts of different kinds of productive causal mechanisms, and so information must be 'cashed out' in terms of the content of the states that direct or reflect the entities and activities of production. But this ascription of content has 'reach"—it extends and unifies the claims made about the whole of the mechanisms and not just parts. It unifies by ascribing a purpose (function) to the whole.

The ascription of information flow to a system is a description based on an interpretationist stance that a researcher takes toward the system. We describe some stages among those constituting the mechanism contain "information" or content that describes the relation between that stage and the initial stage or the end stage. Mechanistic information is information *about* and/or information *for*. In some cases the information about the initial stage is information for or used by the penultimate stage to bring about an appropriate response to what the information is about. So for example, if a leech is prodded by a sharp stimulus, it bends away from that stimulus (Churchland and Senjowski). Or differently said, the leech exhibits an aversive reaction. We may say here that the leech system carries information about the stimulus (and about a state of the world) that causes it to produce an aversive behavior.

Calling the response an aversive behavior connects it back to the initiating noxious or dangerous environmental event, the prodding.

Note that the behavior caused by the workings of the mechanism is not the whole of which the mechanism is a part. The relevant whole may be the organism or in the leech case the organism functioning in an environment.

Information transmission is a way of exhibiting by means of system description the unity that obtains within a system or why is it identified as one system or isolating what is to be identified as important for the system. In all cases the information transmission is a process that connects upstream entities and activities to downstream results. Generally when information as content is present we say the content descriptions shows how the mechanism serves a purpose, or performs a function, which is usually conceived as guiding the bringing about or producing, or contributing to the production of the result completed at the end state, and this guidance is identified as significant or important for the organism or larger containing system. So a cockroach floating in a drink provides information to an adult, but not to many children, for a response of disgust (http://www.apa.org/monitor/oct03/gross.html). This information is somehow carried through the anterior insula as s crucial part of the mechanism. But the end state here is a functional teleological purposive state of the adult human in a real world context.

With mechanisms without information we say the end state is produced by the earlier stages in the mechanisms, and this mechanism or system is unified or counted as one because of the productive activities that provide continuity among the stages. But we would not count the end state as a goal state, since there is no sense in which the initial or subsequent stages are *for the sake of* the end. Of course, we may trivialize information for and about, and use these terms to describe any mechanism, but then this leaves mechanical information unexplained.

The end state becomes a goal state, when the end state is evaluated as being of sufficient importance for the organism, or sometimes for a subsystem of the organism, to warrant the claim that the mechanism works for this purpose. This warrant may come from another theory, e.g. natural selection in evolution, metabolic or homeostatic needs of an organism, etc. It may also come from a perspectival attribution of importance by the researcher, e.g. memory loss in Altzheimer's is a fact, but it is considered important because researchers want to find a cure or a way of retarding its onset.

5. Conclusion

We are now back almost where we started with productive causality. But there is no time to explore this aspect.

References

Achinstein, Peter, (1983), *The Nature of Explanation*, New York: Oxford University Press.

Alexander, Peter, (1963), *Sensationalism and Scientific Explanation*, London: Routledge and Kegan Paul.

Bechtel, William and Robert Richardson (1993) *Discovering Complexity: Decomposition and Localization As Strategies in Scientific Research*, Princeton University Press.
Bechtel, William and Adele Abrahamsen (2005) "Explanation: a mechanist alternative" Studies in History and Philosophy of Biology and Biomedical Sciences, 36, 421-441.

Bechtel, William (2008) *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*, Routledge.

Bechtel, William and Adele Abrahamsen, (2006) "Phenomena and Mechanisms: Putting the Symbolic, Connectionist, and Dynamical Systems Debate in Broader Perspective" in R. Stainton, ed. *Contemporary Debates in Cognitive Science*, Oxford: basil Blackwell.

Jim Bogen and Peter Machamer (2006-7) "Information(s), Mechanisms and Explanations" (draft mss.).

Chafe, W. L. (1970) *Meaning and the structure of language*, University Press, Chicago,

Churchland, Patricia and Terrance Senjowski (1992) *The Computational Brain*, MIT

Encyclopedia of Linguistics

Craver, Carl and William Bechtel (2007)

Craver, Carl F. (2007) *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. New York: Oxford University Press.

Fehr, Carla (forthcoming) "Explanations of the evolution of sex: A plurality of local mechanisms", Minnesota Series in the Philosophy of Science.

Friedman, Michael (1974) 'Explanation and Scientific Understanding" Journal of Philosophy LXXI, 5-19

Genter, Deidre (1978) "On Relational Meaning: The Acquisition of Verb Meaning" Child Development 49, 988-998.

Hall, Ned (2004) "Two Concepts of Causation" in J. Collins, N. Hall, and L. A. Paul, eds., *Causation and Counterfactuals*. Massachusetts: The M. I. T. Press. (http://socrates.berkeley.edu/~fitelson/269/Hall_TCOC.pdf)

Hanson, Norwood Russell (1958) *Patterns of Discovery*, Cambridge: Cambridge University Press

Harré, Rom, (1970) *The Principles of Scientific Thinking*, London: Macmillan.

Hempel C.G and P. Oppenheim (1948) "The Logic of Scientific Explanation" reprinted in C. G. Hempel, *Aspects of Scientific Explanation*, Free Press.

Kitcher, Philip (1999) **"**Unification as a Regulative Ideal*" Perspectives on Science* - Volume 7, Number 3, pp. 337-348

-------, (1989) "Explanatory Unification and the Causal Structure of the World," *Minnesota Studies in the Philosophy of Science*, XIII.

--------, (1981) "Explanatory Unification" *Philosophy of Science*, 48, 507-531.

--------, (1976) Explanation, Conjunction and Unification" Journal of Philosophy LXIII, Mackie, J.L. (1980) *The Cement of the Universe,* Oxford Clarendon Press.

Mandler, Jean (2006) "Actions organize The Infant's World" in K. Hirsh-Pasek & R. Michnick Golinkff, eds. *Action meets Word: How Children Learn Verbs*, Oxford UP.

Millikan, Ruth Garrett (2004) *Varieties of Meaning: The 2002 Jean Nicod Lectures*, MIT.

Mitchell, Sandra (2009) *Unsimple Truths: Science, Complexity and Policy*. University of Chicago Press.

Mitchell, Sandra (2003) *Biological Complexity and Integrative Pluralism*. Cambridge University Press.

Roach, John (2001) "Fear of Snakes, Spiders Rooted in Evolution, Study Finds" National Geographic News, October 4, 2001.

Ruddick, William (1968) "Causal Connection" *Synthese* 18, 46-67.

Stegman, Ulrich E (2006) "Genetic Information as Instructional Content" *Philosophy of Science*, 72, 425-443

Strevens, Michael*, Depth: An Account of Scientific Explanation*, Harvard University Press, 2008.

Strevens, Michael (2007) "Why Represent Causal Relations" to appear in A. Gopnik and I. Schultz, eds. *Causal Learning, Psychology, Philosophy and Computation*, Oxford

Michael Tomasello (1999) The cultural Origins of Human Cognition, Harvard UP, Chapter 5: "Linguistic Constructions and Even Cognition". 134-160

Toulmin, Stephen (1953) *Philosophy of Science*, Harper and Row.

Woodward (2007) mss. "Causal Specificity"

Woodward, James (2003) *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.