Rethinking Mechanistic Explanation[*]

Stuart Glennan

Department of Philosophy and Religious Studies

Butler University

12/21/00

On-line Version

Abstract

Philosophers of science principally associate the causal-mechanical view of scientific explanation with the work of Peter Railton and Wesley Salmon. In this paper I shall argue that the defects of this view arise from an inadequate analysis of the concept of mechanism. I contrast Salmon's account of mechanisms in terms of the causal-nexus with my own account of mechanisms, in which mechanisms are viewed as complex systems. After describing these two concepts of mechanistic explanation, I show how the complex-systems approach to mechanism avoids certain objections that can be raised against Salmon's version of the causal-mechanical approach. I conclude by discussing how mechanistic explanations in the complex systems approach can provide understanding by unification.

1 Introduction

Wesley Salmon (1989) and Phillip Kitcher (1989) both claim that philosophical work on scientific explanation has been dominated by two traditions, which Kitcher terms the "top-down" and "bottom-up" traditions. The top-down approach explains an event by showing it to be part of a larger nomological or explanatory pattern, while the bottom-up approach explains an event by describing the network of causes that are efficacious in bringing that event about. The top-down approach originates with Hempel and Oppenheim's (1948) deductive-nomological model and is most prominently

[*] Thanks to Lindley Darden for helpful comments on an earlier draft of this paper.

represented in the contemporary debate by Kitcher's theory of explanatory unification (1981, 1989). The bottom-up approach may fairly be said to originate with Scriven's critique of the DN model, but it is now represented most clearly in the causal-mechanical approach developed by Railton (1978, 1981) and Salmon (1984, 1994).

While Salmon and Kitcher both acknowledge that many explanations unify and that many other explanations exhibit causal mechanisms, they differ sharply on which of these approaches is fundamental. Salmon claims that unification occurs because many physical processes utilize the same basic causal mechanisms, while Kitcher suggests that our claims about causes are ultimately based upon unifying explanatory patterns. As Kitcher puts it, "the unification view … proposes to ground causal claims in claims about explanatory dependency rather than visa versa" (Kitcher 1989, 436).

My own view is that Salmon's position is the correct one: it is causes that underwrite explanations rather than explanations underwriting causes. But while this is my ultimate position, the criticisms raised of Salmon's account by Kitcher (1989), Woodward (1989), Hitchcock (1995) and others are undoubtedly strong. Salmon's problem lies not with his advocacy of the causal-mechanical approach, but rather with his understanding of the nature of mechanisms. In particular, Salmon's account, which explicates causal mechanisms in terms of interactions between causal processes, fails to elucidate important causal and explanatory features of mechanisms. A more adequate account of the nature of mechanisms clarifies how mechanistic explanation can lead to unification.

I begin with a discussion of the concept of mechanism implicit in Salmon's and Railton's theories of explanation. I then describe an alternative approach of mechanisms that I call the complex-systems approach. Finally, I show how this approach leads to a theory of explanation that avoids many of the pitfalls of the Railton/Salmon view, and also how it provides insight into explanatory unification.

2 Salmon and Railton on Causal-Mechanical Explanation

Peter Railton (1978) was the first to introduce the idea of mechanism into the contemporary literature on explanation. His deductive nomothetic model of probabilistic explanation (the DNP model) was meant to be an alternative to Hempel's inductive statistical (IS) model. Railton was concerned with Hempel's requirement that the explanans of an IS explanation render the explanandum probable or nomically expectable. On Railton's view, explanations describe causes, and sometimes the sequence of events leading up to the event to be explained may be improbable. According to Railton, while an explanation of some event may include a reference to a law that renders the event nomically expectable, the account must be supplemented by "an account of the mechanism(s) at work" (1978, 748). Railton is deliberately vague on just what a mechanism is, indicating only that an "account of the mechanism(s)" is "a more or less complete filling-in of the links in the causal chains" (ibid.).

Salmon's work on causal-mechanical explanation, beginning with his seminal *Scientific Explanation and the Causal Structure of the World* (1984), can be seen in large part as an attempt to elaborate on Railton's earlier account of mechanistic explanation. Curiously, nowhere in this book (or to my knowledge, elsewhere in his work) does Salmon offer an explicit definition of mechanisms. Rather, he provides a characterization of what he calls the "causal nexus", which he takes to be a vast network of interacting causal processes. Much of Salmon's work is directed at providing an adequate definition of just what counts as a causal process and a causal interaction. According to his definition, a process is an entity that maintains a persistent structure through space-time, a causal process is a process capable of transmitting changes in its structure, and a causal interaction is an intersection between causal processes in which an alteration of the persistent properties of those processes occurs. In Salmon's original formulation interactions were defined in terms of a counterfactual criterion of mark transmission, but in more recent versions (1994) he has tried to eliminate the reference to counterfactuals, relying instead upon the concept of a conserved quantity to define causal interactions. While Salmon has spent considerable effort subsequent to 1984 trying to refine his characterizations of processes and interactions, his view of the causal nexus as a network of interacting processes remains essentially unchanged.

Causal-mechanical explanation exemplifies what Salmon has dubbed the ontic conception of explanation. An explanation is not to be understood as an argument (a linguistic entity) but as the description of certain features of a mind-independent reality—the causal structure of the world. It is notable that Salmon uses a definite article. There is such a thing as *the* causal structure of the world. While epistemic and pragmatic considerations will lead any description the causal structure of the world to be fragmentary, this description is explanatory in virtue of its reference to this mind-independent causal structure.

To clarify the relationship between ideal and actual explanations, Railton has introduced his concept of the "ideal explanatory text". The ideal explanatory text would consist of

> an interconnected series of law-based accounts of all the nodes and links in the causal network culminating in the explanandum, complete with a fully detailed description of the causal mechanisms involved and theoretical derivations of all the covering laws involved. This full-blown causal account would extend, via various relations of reduction and supervenience, to all levels of analysis (Railton 1981, 246-47).

Note here the in-principle reductive character of causal structure. While it is granted that higher levels of analysis may be essential to the practical business of explanation, the real causal structure of the world is microphysical. This is an assumption that is questioned by the complex-systems approach.

3 The Complex-Systems Approach to Mechanism

While the term "causal-mechanical explanation" has come in the literature to refer to the theories of explanation proposed by Railton and Salmon, a number of philosophers have been developing a quite different analysis of mechanism—one that leads to an alternative conception of causal-mechanical explanation. Central to this alternative conception is the idea that a mechanism is a kind of "complex system." The germ of this conception can be found in the work of Wimsatt (1974), which in turn has its antecedents in the work of Herbert Simon. This conception has been developed to a much greater degree in my own work (Glennan 1992, 1996) as well as that of Bechtel and Richardson (1993) and, most recently, Machamer, Darden and Craver (2000). While there are differences between these approaches, they all share the view that mechanisms are complex systems.

A number of definitions of mechanism have been suggested, but my preferred one is as follows:

> (M) A mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations.

I shall not repeat arguments I have made elsewhere (Glennan 1992, 1996) for this definition, but a few clarifications are in order. First, mechanisms are not mechanisms *simpliciter*, but mechanisms *for* behaviors. The same complex system may exhibit several different behaviors, and the way in which the system is decomposed into parts will depend upon which behavior is under consideration. To take a well-worn example, a heart is both a mechanism that pumps blood and a mechanism that produces noise, and the heart qua pump may admit of a different decomposition than the heart qua noisemaker. Second, mechanisms consist of a number of parts. These parts must be objects, in the most general sense of that term. They must have a relatively high degree of robustness or stability; that is, in the absence of interventions, their properties must remain relatively stable. Generally, but not always, these parts can be spatially localized.

A mechanism operates by the interaction of parts. An interaction is an occasion on which a change in a property of one part brings about a change in a property of another part. For instance, a change in the position of one gear within a clock mechanism may bring about the change in the position of an interlocking gear. ‘Interaction’ is a causal notion that must be understood in terms of the truth of certain counterfactuals. The stipulation that these interactions can be characterized by invariant, change-relating generalizations is meant to capture the relevant counterfactual truth claims.

I borrow the phrase ‘invariant, change-relating generalization’ from Jim Woodward’s recent and insightful paper on explanation and invariance (Woodward 2000). To say that a generalization is change-relating is to say that it describes a relationship between two (or perhaps more) variables in which an intervention that changes one variable will bring about a change in another variable. For instance, the Boyle-Charles law can be regarded as a change-relating generalization, since intervention on one variable, say the temperature of an enclosed volume of gas, can bring about a change in another variable, the pressure of the gas. Not all true counterfactual-supporting generalizations are change-relating. For instance (to take one of Woodward’s examples), a counterfactual-supporting generalization describing the relationship between falling barometers and storms is not change-relating, because an intervention tchange the position of the barometer needle will not bring about a change in the weather. Similarly, conservation laws and other laws that characterize constraints on possible states of physical systems—what advocates of the semantic conception call 'laws of coexistence'—are not change-relating. Change-relating generalizations bear a close affinity to what Nagel (1979) called "causal-laws" and what Hempel (1965) and others have called "laws of succession".

The reason that I have avoided calling these generalizations causal laws or laws of succession is that they lack a property philosophers commonly attribute to natural laws—namely, exceptionless universality. While these generalizations are universal in form, they are not exceptionless. Invariance admits of degrees, and invariant generalizations may fail to hold under extreme values of the variables that figure in the generalization or under certain changes in background conditions not mentioned in the generalization itself (*cf.* Woodward 2000, §3). In not calling such generalizations laws, I have conceded to prevailing philosophical usage, but I should note that many generalizations that scientists call laws (like the Boyle-Charles law or Mendel's law of segregation) are invariant but do not qualify as laws in the philosopher's sense.[1]

To use invariant change-relating generalizations to characterize interactions, we must add the additional stipulation that the generalizations relate *direct* changes. This stipulation is intended to rule out generalizations that would truly describe relationships between two parts in which one part indirectly caused a change in the properties of a second part by changing the properties of one or more intervening parts. For instance, it is intended to rule out a generalization that relates a change in the position of one gear to a change in the position of a second gear, when this change is accomplished by the first gear turning a third gear which is connected in turn to the second gear.

An important feature of the complex-systems approach to mechanisms is that it is hierarchical. The objects that are parts of mechanisms may themselves be complex mechanisms that can be decomposed into further parts. For instance, a description of the mechanisms responsible for the distribution of oxygen through a human body may decompose the complex system (the body) into parts that include the heart, the lungs, arteries, veins, capillaries, and blood cells. But the heart (and for that matter, every other part) is itself a complex system whose capacities can be explained by decomposition of its parts.

---

[1] In earlier papers (1992, 1996) I have in fact called such generalizations 'laws' with the caveat that these laws must be understood in a more homely way than philosophers typically understand them. Woodward

The hierarchical character of mechanisms has implications for the idea of a direct interaction. What is meant by a direct interaction will vary with the level of decomposition. For instance, at the first level of analysis, we may treat the lungs as "black boxes" that oxygenate blood passing through them. At a more detailed level, we can describe the details of the structure of the lung and the intertwined capillary system, and at a still more detailed level, the intracellular structures that allow hemoglobin to transport oxygen. What counts as a direct interaction between parts depends upon what the parts are—i.e., upon the level of decomposition.[2]

There are two important respects in which the Salmon/Railton account differs from the complex-systems account. Perhaps the most immediate difference is that Salmon/Railton mechanisms are *sequences of interconnected events* while complex-systems mechanisms are *things* (or objects). When Salmon or Railton talk about mechanisms, they are generally talking about a chain or web of events leading up to a particular event. For instance, they might speak about the chain of events that lead to the breaking of a certain window: a boy hit a baseball; the baseball ricocheted off the tree and crashed into the window. While the sequence of events leading to the breaking of the window certainly involves some entities that are stable enough to be called objects—the

---

and others have persuaded me to bow to prevailing philosophical usage.

[2] The reader may note that the analysis of a complex mechanism resembles the strategy Cummins (1975) calls 'functional analysis'. While I think that in a number of respects the explanatory strategy associated

boy, the baseball, the bat, the tree, and the window—the complex of the boy, bat, baseball, tree and window do not form a stable enough configuration to be called a thing or object.

Compare this to examples of mechanisms on the complex-systems conception: for instance, watches, cells, organisms, and social groups. These mechanisms are systems consisting of relatively stable arrangements of parts. In virtue of these stable arrangements, the system as a whole has a stable set of dispositions—the behaviors of these mechanisms. These dispositions can manifest themselves at more than one time and place. In this sense, the behavior of a complex-systems mechanism is general. The watch, for instance, continues its periodic motion more or less indefinitely. Complex-systems mechanisms are general in a second sense. Although any particular mechanism will occupy a particular region of space-time, it is clearly an important feature of our world that it often contains many tokens of a single type of mechanism. For instance, the human central nervous system contains around a trillion neurons. There are lots of human beings, as well as lots of other organisms that have neurons whose cellular structure is similar to that of human neurons. Consequently, one can develop a general model of neurons that subsumes countless neural events.

---

with mechanistic analysis has more to do with Cummins' approach than with Salmon's and Railton's, there are differences. Craver (2001) explores these, so I will not do so here.

The second important difference between the Salmon/Railton and complex-systems conceptions of mechanism has to do with the treatment of interactions. At a very general level, there is a similarity between the accounts. The parts referred to in (M) are objects, and objects can be understood as causal processes in Salmon's sense. Interactions between these parts will then be intersections in causal processes that introduce changes to the persistent structure of those processes, i.e., changes to the properties of the parts. However, there is a decided difference between the explication of interaction in my and Salmon's accounts. While I have characterized interactions by saying that they are described by direct invariant change-relating generalizations, Salmon has recently (1994) characterized them in terms of exchange of conserved quantities. For instance, two colliding particles interact when they exchange a conserved quantity like momentum. Salmon moved to the conserved quantity approach to avoid certain objections to the counterfactual approach (Kitcher 1989), and while the move has the advantage of characterizing interactions in terms of physical theory rather than the semantics of counterfactuals, it has the disadvantage that it obscures similarities between kinds of interactions among higher-level entities. Consider, for instance, a social mechanism whereby information is disseminated through a phone-calling chain. The people in the chain are the mechanism's parts, and each phone call represents an interaction between parts of the system. Now each such interaction must, in virtue of the supervenience of everything upon physics, involve a string of physical events in which conserved quantities are exchanged. However, to say that my phone call involves the exchange of a conserved quantity is utterly uninformative about the nature of particular interactions over the phone. To talk in terms of exchange of conserved quantities requires us to treat mechanisms at a level at which such talk is intelligible, namely at the level of physics. But different tokens of a single of higher level event type (e.g., a phone call event type) may have nothing in common in terms of their physical descriptions. Thus, if interactions are characterized in terms of exchange of conserved quantities, tokens of higher level interactions cannot be recognized as forming higher-level kinds.

This argument obviously borrows heavily from Putnam's (1973) and Fodor's (1974) arguments for the explanatory autonomy of the special sciences. I am just noting that the parts of mechanisms and their interactions are autonomous in the same sense. Salmon's characterization of interactions in terms of the conserved quantity theory further exacerbates a tendency to treat causal-mechanical explanations as, ideally speaking, micro-physical explanations. This approach prevents Salmon's analysis from providing any insight into the operation of higher level mechanisms.

4 Mechanistic Explanation on the Complex-Systems Approach

Let us consider how to develop an account of mechanistic explanation consistent with the complex-systems approach. Philosophers have generally agreed that scientific explanations come in (at least) two varieties—explanation of singular events and explanation of general regularities. Let us begin with the mechanistic explanation of regularities. Roughly, the idea is this: To mechanistically explain a regularity, one describes a mechanism whose behavior is characterized by that regularity. For instance, a mechanistic explanation of Mendel's law of segregation will describe the meiotic mechanism that produces gametes and show how this mechanism creates (assuming certain background conditions) an equal number of gametes containing each allele of a given locus.

To spell this account out in more detail, we need to introduce another notion—the notion of a mechanical model:

> (MM) A mechanical model is a description of a mechanism, including (i) a description of the mechanism's behavior; and (ii) a description of the mechanism which accounts for that behavior. (Glennan under review)

The two-part characterization of a mechanism, in terms of a behavior and the mechanism that produces it, leads naturally to a two-part characterization of a mechanical model. The behavioral description is a description of the external behavior of a mechanism. The mechanical description is a description of the internal structure—the guts of the mechanism. The distinction between behavioral and mechanical descriptions is roughly the distinction between what a system is doing and how it is doing it.

It is probably not possible to provide an illuminating and general account of the logical form of behavioral descriptions. For the purpose of illustrating how mechanisms can explain regularities, let us confine ourselves to a particular kind of behavior—the behavior of what I call an input-output mechanism. An input-output mechanism is a complex system that is situated in its environment in such a way that there are characteristic environmental events (inputs) that trigger a sequence—perhaps multi-stranded—of interactions between parts of the mechanism. This sequence concludes with some terminating event, the output. A familiar example of an input-output mechanism is a computer running a word processing program: depression of keys on the keyboard are input events while the appearance of characters on the screen are output events. Another input-output mechanism, to borrow an example from Machamer, Darden, and Craver (2000), is the mechanism by which neural signals are transmitted across synapses. The input is a signal coming along the axon of a pre-synaptic neuron, and the output is a signal that is passed from the dendrites of a post-synaptic neuron towards the cell body.

The behavior of input-output mechanisms can be characterized by the same type of invariant change-relating generalizations that characterize interactions between parts of mechanisms. Take such a generalization as an explanandum. Evidently, to the extent that a description of the internal workings of the mechanism elucidates the process by which the parts of the mechanism, triggered by the input event, interact in order to produce the output event, that description can be construed as the explanans. To put it more succinctly, the mechanical description associated with the mechanical model illustrates why the behavioral description is true. It describes the causal process underlying the behavior to be explained.

While many mechanisms are of the input-output variety, there are other kinds of mechanisms whose behaviors are not described in terms of inputs and outputs. Another important kind consists of mechanisms that are responsible for the maintenance of continuous states of affairs or of periodic behaviors. Consider for instance the mechanistic explanation of Kepler's first law of motion. The behavior in question is the elliptical orbit of a planet around a central body. The behavior is explained by the constant gravitational interaction of the two bodies in accordance with Newton's laws.[3] Another important class of mechanism is the class of homeostatic mechanisms. Homeostatic mechanisms, e.g., the mechanisms that regulate temperature in warm blooded animals or the heating system that maintains a constant temperature in my house, maintain some property of a system in the face of perturbations from the environment. Here, the behavior in question is, so to speak, an input-no-output generalization. But, as in the other cases discussed, we can analyze the system into parts (for instance, thermostat, furnace, fan, ducts), and indicate how the interactions of these parts are responsible for the behavior in question—here the maintenance of a (nearly) steady temperature.

[3] I choose this example in part to make clear the possibility of providing a mechanistic explanation of phenomena that are in no sense designed. The example, however, raises an important question— namely, whether parts of mechanisms can act at a distance. My view is that nothing in the concept of mechanism rules out this possibility a priori. Einstein's approach to gravitation does not require action at a distance, but, in this like other cases, we must accept the principle that what counts as an interaction must be decided by the various scientific disciplines, not by metaphysical or conceptual analysis. This case, as well as the

The process of mechanistic explanation requires one to formulate, perhaps in very sketchy terms, a mechanical model. The two parts of this model, the behavioral and the mechanical descriptions, are respectively the explanandum and explanans in the mechanistic explanation. Because these descriptions are generally statements, the mechanical model can be construed as an argument in which the statements in the mechanical description are premises and the conjunction of statements in the behavioral description is the conclusion. So construed, explanations apparently are arguments, just as Hempel and Kitcher suggest. I would like to resist this conclusion. While it is sometimes the case that a description of the parts of the mechanism will entail a description of the mechanism's outward behavior, the explanation lies not in the logical relationship between these descriptions but in the causal relationships between the parts of the mechanism that produce the behavior that is described. Sometimes we can construct models in which logical relationships between semantic entities mirror causal relationships between physical entities, but the logical relationship is not required. For one thing, it is possible to provide mechanical models using descriptive tools, like diagrams and flow charts, for which entailment relationships are not defined. For another, causal relationships, unlike entailment relationships, can be broken by extreme values in variables or by changes in background conditions.

---

general question of the relationship between empirical and conceptual constraints on mechanisms is discussed more fully in Glennan 1992 and 1996.

The mechanisms so far discussed in this section are complex systems consisting of relatively stable configurations of parts that give rise to robust behaviors that can be expressed by invariant generalizations. I have argued elsewhere (1996, 1998) that most of the generalizations that scientists call laws are in fact descriptions of the behavior of mechanisms. Mendel's laws describe aspects of the behavior of mechanisms that transmit genetic material; Kepler's laws describe aspects of the behavior of gravitational mechanisms, and so on. Laws of this kind I call mechanically explicable laws. While most laws are mechanically explicable, inevitably there must be some laws that are not. For instance, given our present understanding of electricity and magnetism, it appears that Maxwell's equations are not mechanically explicable. There is not, for instance, a mechanical ether consisting of particles whose interactions could explain the propagation of electromagnetic waves. Laws such as these, which I call fundamental laws, represent the brute nomological facts of our universe.

Whether we use the term 'law' to refer to an invariant generalization that describes the behavior of a mechanism depends upon a number of factors. These include the ubiquity of the mechanism, the degree to which its behavior is invariant under changes in background conditions, and the relevance of the generalization to explaining other phenomena of interest to us. We call Kepler's laws 'laws' because, for a variety of reasons primarily of a cultural sort, the problem of predicting the position of planets has been taken to be a major problem of astronomy. We do not feel the same way about generalizations about the behavior of Coke machines, but, from a empirical or conceptual point of view, the only significant difference between them is that the behavior of planets is more reliable than the behavior of Coke machines.

Mechanisms consisting of complex systems with stable relationships between parts can explain many of the regularities that occur in the world. But what of the explanation of particular events? I shall argue that there are two different sorts of events that we might seek to explain. On the one hand, there are events that are the product of the operation of reliable mechanisms. On the other, there are events that are not the product of such mechanisms. These latter events I shall call “genuinely singular events”.

Explanation of events that are the product of reliable mechanisms follow a pattern familiar from standard accounts of explanation by nomic subsumption. Consider, for instance, the explanation of the fact that my son John has blue eyes. Blue eyes involve expression of a recessive gene.[4] Given that my wife has blue eyes and I do not, she must carry two copies of the recessive allele while I carry one. Consequently, given the mechanisms of gamete formation, reproduction, and expression of genes determining eye color, there is a probability of 50% (relative to a generally satisfied set of background conditions) that John would have blue eyes, as in fact he does. Thus our explanation looks like this:

My wife is homozygous for the recessive allele for blue eyes.
I am heterozygous for the recessive allele for blue eyes.
John is our biological child.
If a child has one parent who is homozygous for a recessive allele and one who is heterozygous, the probability of the child being homozygous is 50%
John has blue eyes. [p = .5]

This is a typical example of explanation by nomic subsumption (except that it would not meet Hempel's high probability requirement). Its only connection with mechanistic explanation is that the statistical "law" which serves as its fourth premise is mechanically explicable.

While arguments of this kind are explanatory, it is also possible to ask for a different kind of explanation of the event in question. Rather than show how a certain mechanically explicable law gives a certain event some degree of probability, it is possible to explain this event by going through a play-by-play description of the operation of the relevant mechanism on that occasion. In this case, one could tell the story of the formation of the egg and sperm, the events leading to the particular sperm fertilizing the particular egg, and the subsequent developmental sequence leading to the expression of the particular gene combination in John's blue eyes. While explanations of this kind are possible, it is generally preferable with events of this sort to pursue a two-part explanatory strategy. First, provide a nomic subsumption explanation of the event,

[4] Although this case was once considered a classic case of an effect produced by a recessive gene at a single locus, there is now evidence that this trait is polygenic. In most contexts, however, the trait exhibits the pattern described, and the example will serve my purposes.

and, second, provide a mechanistic explanation of the law or invariant regularity under which the event is subsumed. Such a two part strategy explains both what caused the particular event in question and how that cause fits into a common pattern.

Kim Sterelny (1996) has offered some useful terms to characterize the difference between the two types of explanations of a particular event like my son's developing blue eyes. Sterelny suggests that, depending on our explanatory interests, we can explain a particular event either as the result of a robust process or as the result of an actual sequence. To illustrate this claim, he considers two different explanations of the outbreak of World War I (1996, 195). The actual-sequence explanation describes in detail the events leading to declarations of war in August 1914—the visit of the Archduke to Sarajevo, his assassination, the Austrian response and Russian responses, and so on. The robust-process explanation offers a more general description of system of alliances, mobilization plans, nationalism, militarism and other features of Europe in 1914 that in retrospect suggest that World War I was more or less bound to happen. Sterelny argues that these explanations are both correct, but are appropriate answers to different explanatory questions.

Although his position is not completely clear, Sterelny seems to suggest that we can explain any event either from the robust process or actual sequence perspective. However, if, as I have suggested, there is a class of genuinely singular events that are not the product of the operation of reliable mechanisms, then such events could not be given a robust-process explanation. Consider, for instance, an explanation of how I met my wife. That event, like many events in which people meet, involved a confluence of events that were not to be expected and will not be repeated. In our case, for instance, an explanation of the event would, at the very least, involve an explanation of why each of us chose to attend graduate school at the University of Chicago, an explanation of the factors that led us to apply for student housing, and an explanation of the process that led the housing office to place each of us in the same building. This is an actual sequence explanation *par excellence*, but there does not appear to be a corresponding robust process.

Given that robust processes arise from the operation of reliable mechanisms, it might seem that events that are not the product of robust processes cannot be given a mechanistic explanation. Here, however, it is useful to recall the distinction made above between mechanisms as things and mechanisms as processes. The actual sequence of events is a process; it is simply not a *robust* process. Mechanisms in the sense of Salmon and Railton can characterize the actual sequence of events and connecting processes that bring about a particular explanandum event, even if that sequence does not represent the operation of a robust process. Such an explanation can be called a fragile actual-sequence explanation.

Even though mechanisms as complex systems cannot explain fragile actual sequences, there is still one point where the Salmon/Railton account can be improved upon by reference to the complex-systems conception of mechanism. This point concerns how an "actual sequence of events" is specified. Because events in the universe are capable of all sorts of descriptions at various levels of spatial and temporal grain, it is not exactly clear what could be meant by the "actual sequence of events". In the case of World War I, what events are in this actual sequence? Sterelny suggests that this sequence includes such things as the firing of a pistol at the Archduke, but why do we describe events at this level? Why not, for instance, describe the trajectories of molecules in the Archduke's body?

I think that even in fragile actual-sequence explanations, the description of a process is guided by considerations of robustness and reliability. The "actual sequence" of events leading up to the beginning of the first World War could aptly be described as a *fragile process with robust parts*. This description is apt because, even in actual-sequence explanations, we characterize processes in terms of interactions of robust entities. Gavril Princip, the Archduke, his driver, etc. are all robust entities whose interactions comprise the sequence of events leading to the First World War. The particular arrangement of the parts (e.g., the location of the Archduke and the Assassin on the fateful day) may be ephemeral, but the parts themselves (the Archduke and the Assassin) are not. Consequently, it was predictable what was going to happen when one part, the Assassin, used another part, the Gun, at close range to shoot at a third part, the Archduke. Much the same thing can be said about my earlier example of the ball breaking the window. While the particular configuration of ball, bat, tree and window was not robust, those things themselves are robust. In both cases, it is the robustness of these parts that accounts for our decision to describe actual sequences in terms of interactions between them.[5]

5 Unification by Common Mechanisms

Whatever the merits of their particular proposals, Kitcher, Friedman and others who have advocated unification models are right to insist that scientific explanations generally afford us with a more unified understanding of the world. How can the proponent of causal-mechanical explanation account for this property of explanation? Salmon makes the following proposal:

> We explain events by showing how they fit into the causal nexus. Since there seem to be a small number of fundamental causal mechanisms, and some extremely comprehensive laws that govern them, the ontic conception has as much right as the epistemic conception to take the unification of natural phenomena as a basic aspect of our comprehension of the world. *The unity lies in the pervasiveness of the underlying mechanisms* upon which we depend for explanation. (Salmon 1984, 276).

In concluding this paper I would like to explore to the extent to which Salmon's proposal does justice to the idea of explanation as unification.

When Salmon speaks of "fundamental causal mechanisms" governed by "extremely comprehensive laws", he seems to have in mind some set of basic types of

[5] Hitchcock's (1995) criticism of Salmon's account of explanation seems to me to be effective precisely because of Salmon's tendency to overemphasize actual sequences at the expense of robust processes. Hitchcock argues that Salmon's account of the causal nexus leaves out an essential modal component—one that refers to capacities of the entities involved in an explanation. Robust processes or even fragile processes with robust parts provide this missing modal component.

microphysical interactions, governed by what I have termed fundamental physical laws. Salmon's view is that mechanistic explanations provide unifying understanding of the world by appealing to these laws. This view is consistent with some of the best examples of explanatory unification. Consider, for example, Newton's explanation of the motion of the planets, the behavior of projectiles, and the changing tides. In showing that all of these phenomena can be explained in terms of the operation of a single fundamental force, gravity, Newton engaged in an impressive act of explanatory unification.

While there are other examples in which we appeal to fundamental laws to explain a variety of apparently diverse phenomena, I would argue that there are many kinds of unification that do not involve appeal to *fundamental* mechanisms or laws. Our discussion in the preceding section showed how characterization of the mechanisms that give rise to robust processes often takes one away from the microphysical detail of actual sequences. Robust higher-level mechanisms are, to use Wimsatt's (1994) phrase, dynamically autonomous—that is, the higher level processes are relatively insensitive to changes in the microphysical processes on which they supervene.

The complex-systems approach to mechanisms does not suppose that unification derives from unity of fundamental mechanisms. According to the complex-systems approach, mechanisms are collections of parts and parts are objects, but the objects that are parts of mechanisms may themselves be complex structures. An information processing mechanism used in human cognition may have interacting parts in the sense required by the complex-systems conception, even if these parts are spatially distributed and realized by different neurological components in different instances of the mechanism. This and other examples show that the explanatory unification afforded by the mechanistic approach derives not only from the commonality of fundamental laws but from the existence of mechanisms that have a common higher-level structure even if they differ in microstructure.

Rob Skipper has recently (1999) pursued the idea that explanatory unification can be pursued in this way, using as his example what Darden and Cain (1989) have called selection-type theories. Skipper shows how selection processes in different domains (evolutionary biology, immunology and neurobiology) are unified by what he calls a mechanism schema. According to Skipper, selection processes in each of these domains involve the operation of a common kind of mechanism—a selection mechanism. These processes are obviously heterogeneous from a microphysical point of view, but they share a common higher-level structure; they all are processes in which a certain trait becomes entrenched within a population because that trait confers benefits on individuals in the population who bear that trait.

Skipper recognizes that there is considerable similarity between his approach of unification through common mechanism schemata and Kitcher's approach of unification through common argument schemata. Skipper's major claim is that Kitcher cannot accept a "general selection schema"—an argument schema that would unify selection across domains—because it would not meet the “stringency requirements” that Kitcher must place on argument schemata in order to bar spurious unification. I will not discuss Skipper’s argument in detail, but only indicate its upshot: derivational relationships between explanans statements and explanandum statements are not the same as causal relationships between explanantia and explananda. Explanatory unification consists in recognizing similarities in causal relationships rather than similarity in derivational patterns.

A large part of the power of mechanistic explanations to unify lies in the fact that scientists can construct mechanical models of differing degrees of abstraction (cf. Glennan under review). Consider as an example, a model of the cell. The simplest models of cells will describe only the gross features common to all cells. More detailed models will differentiate between types of cells—prokaryotic cells versus eukaryotic cells, human cells versus drosophilia cells, germ cells versus somatic cells, etc. The simplest cell models may be adequate to explain certain behaviors of cellular mechanism (like the transference of waste products through cell membranes) while more detailed models will explain more specialized functions. Models of this kind will form a hierarchy, with the most abstract models applying to the widest range of instances. Note that this kind of unification is not available on the Salmon/Railton account of mechanism, since these abstract models will have to abstract away from details of actual sequences.

To summarize, the strategy open to the mechanist is to argue that explanatory unification is achieved by finding pervasive mechanisms that explain apparently diverse phenomena. This is the strategy proposed by Salmon, and endorsed by Skipper. I have added to this account the clarification that unification often occurs by showing how physically diverse processes are nonetheless instances of common higher-level mechanisms. The process of unification is further advanced by developing a hierarchy of related mechanical models of varying degrees of abstraction and generality. Kitcher's insights are incorporated into this picture by suggesting that his argument schemata are often schematic descriptions of mechanisms.

Even if the complex-systems approach can provide a reasonable account of how mechanistic explanations can provide unification, there remains an important metaphysical controversy between the mechanist and the unificationist over the priority of the concepts of explanation and causation. The unificationist argues that explanations precede causes while the mechanist argues that mechanisms precede causes. Kitcher represents the unificationist position as follows:

> The heart of the unification approach is that we cannot make sense of the notion of a basic mechanism apart from a systemization of the world in which as many consequences as possible are traced to the action of as small a number of basic mechanisms as possible. (Kitcher 1989, 497).

The mechanist has a simple answer to this: Kitcher has confused epistemological and metaphysical questions. The claim that we identify basic mechanisms by choosing those mechanisms that maximally unify our understanding of diverse phenomena is really a claim about how we discover and test theories about mechanisms. It is related to the claim that we choose scientific theories on the basis of their simplicity and scope. While there is undoubtedly some truth in Kitcher's claim about how we discover and test theories about fundamental mechanisms, we can still "make sense of the notion of a basic mechanism" apart from this global perspective. In fact, our concept of causation suggests that events occurring at some point in space and time are explained as the consequence of the operation of causal mechanisms operating in that region of space and time. Our global evidence suggests that—quantum mechanics aside—causality is everywhere local.

This brief argument will probably not convince instrumentalists of the reality of causes, but I hope it is enough to show that Kitcher's position is not the only possible interpretation of the global character of theory choice. And if it is possible to treat causes as prior to explanations, it is also possible to see how causal explanations can provide explanatory unification.

References

Bechtel, William and Robert Richardson (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton: Princeton University Press.

Craver, Carl (2001) "Functions, Mechanisms, and Hierarchy: Role Functions in Mechanistic Relief", *Philosophy of Science.*

Cummins, Robert (1975), "Functional Analysis", *Journal of Philosophy* 72: 741-765.

Darden, Lindley and Joseph Cain (1989), "Selection Type Theories", *Philosophy of Science* 56: 106-129.

Fodor, Jerry (1974), "Special Sciences, or The Disunity of Sciences as a Working Hypothesis", *Synthese* 28: 97-115.

Glennan, Stuart (1992), *Mechanisms, Models and Causation.* Dissertation. Chicago: University of Chicago.

______ (1996), "Mechanisms and the Nature of Causation", *Erkenntnis* 44: 49-71.

______ (under review), "A Model of Models".

Hempel, Carl (1965), "Aspects of Scientific Explanation" in *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York: Free Press.

Hempel, Carl. and Paul Oppenheim (1948), "Studies in the Logic of Explanation", *Philosophy of Science* 15: 135-175.

Hitchcock, Christopher (1995), "Discussion: Salmon on Explanatory Relevance", *Philosophy of Science* 62: 305-320.

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

______ (1989), "Explanatory Unification and the Causal Structure of the World", in Kitcher and Salmon 1989, pp. 410-505.

Kitcher, Phillip and Wesley Salmon (eds.) (1989), *Scientific Explanation, Minnesota Studies in the Philosophy of Science*, Vol. XIII. Minneapolis: University of Minnesota Press.

Machamer, Peter, Lindley Darden and Carl Craver (2000), "Thinking about Mechanisms", *Philosophy of Science* 67: 1-25.

Nagel, Ernest (1979) *The Structure of Science: Problems in the Logic of Scientific Explanation*. 2nd ed. Indianapolis, IN: Hackett.

Putnam, Hilary (1973), "Reductionism and the Nature of Psychology", *Cognition* 2: 131-146.

Railton, Peter (1978), “A Deductive-Nomological Model of Probabilistic Explanation”, *Philosophy of Science* 45: 206-226.

______ (1981), “Probability, Explanation, and Information”, *Synthese* 48: 233-256.

Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

______ (1989), “Four Decades of Scientific Explanation”, in Kitcher and Salmon (1989), pp. 3-219.

______ (1994), “Causality without Counterfactuals”, *Philosophy of Science* 61: 297-312.

Skipper, Robert (1999), “Selection and the Extent of Explanatory Unification”, *Philosophy of Science* 66: S196-S209.

Sterelny, Kim (1996), “Explanatory Pluralism in Evolutionary Biology”, *Biology and Philosophy* 11: 193-214.

Wimsatt, William (1974), “Complexity and Organization”, in K. Schaffner and R. Cohen (eds.),*PSA 1972: Boston Studies in the Philosophy of Science* 20, pp. 67-86.

______ (1994), “The Ontology of Complex Systems: Levels, Perspectives and Causal Thickets”, *Canadian Journal of Philosophy* suppl. vol. 20: 207-274.

Woodward, James (1989), “The Causal Mechanical Model of Scientific Explanation”, in Kitcher and Salmon (1989), pp. 357-383.

______ (2000), “Explanation and Invariance in the Special Sciences”, *British Journal for the Philosophy of Science* 51, 197-254.