The heuristics of mechanism discovery: Review of Bechtel and Richardson's *Discovering Complexity*
Phyllis McKay Illari

Bechtel & Richardson's *Discovering Complexity* was originally published in 1993, and reissued in 2010. The original book is an impressive extended examination of the process of discovering mechanistic explanations, illustrated throughout with detailed historical case-studies including fermentation, cellular respiration, phrenology, language processing, developmental genetics, various approaches to the nervous system, and the debate over vitalism. Bechtel & Richardson explain that they use these as sources of insight into features of science that their philosophical account is supposed to explain. The 2010 edition also offers a new preface and useful extended introduction giving Bechtel & Richardson's own reassessment of their work.

Discovering Complexity is rich in both historical detail and philosophical significance. I have a great deal to say about it but – forced to choose – I will examine two of the elements of the work that Bechtel & Richardson themselves identify as important: what are mechanisms, and the heuristics of mechanism discovery. As the philosophical significance of the view might be at risk of being lost in the historical detail, this examination will also serve to emphasize it.

What are mechanisms?

Naturally, *Discovering Complexity* gives Bechtel & Richardson's characterization of a mechanism: 'A machine is a composite of interrelated parts, each performing its own functions, that are combined in such a way that each contributes to producing a behavior of the system. A mechanistic explanation identifies these parts and their organization, showing how the behavior of the machine is a consequence of the parts and their organization.' (p17.) Bechtel's view on mechanisms, developed in collaboration with both Abrahamsen and Richardson, is now well known and widely cited in the mechanisms literature. Even so, there is a great deal of interesting work in *Discovering Complexity* in the historical examination of mechanism and of the empirical work that has shaped the concept.

Because of these, *Discovering Complexity* can be read as an extended answer to an uneasy feeling that the concept of mechanism is somehow deeply problematic. Bechtel & Richardson offer a powerful modern vision of mechanism and the detail of the search for mechanistic explanation which goes a long way towards alleviating any unease.

Bechtel & Richardson first recognize the problem: given a phenomenon, a function, or a task, there are indefinitely many ways to decompose that task into sub-tasks; given a physical system, there are also indefinitely many ways to decompose that system structurally into sub-parts. Thus, any resulting decomposition is too easy. The result is too unconstrained to count as much more than a possible description of the task, or the system.

However, it is not easy to try to map a functional decomposition to a structural decomposition. For Bechtel & Richardson, the core of mechanistic explanation is decomposition and localization. The phenomenon or task is decomposed functionally, the system is decomposed structurally, and scientists attempt to localize sub-tasks in structural parts of the system. The resulting mechanistic explanation is formed from the *mutual constraints* of decomposition and localization. Simultaneously satisfying these mutual constraints, including integrating descriptions of the system at different levels, is not at all

easy, as Bechtel & Richardson illustrate using detailed examples of successful empirical work. Far from being massively underdetermined, building even a single model that satisfies both constraints can be difficult.

So while Bechtel & Richardson's characterization of a mechanism has been noticed, it is perhaps due still more notice. Certainly, even those familiar with it may gain in depth of understanding by reading *Discovering Complexity*.

The heuristics of mechanism discovery

Bechtel & Richardson identify this second element as the most neglected aspect of their work. Yet for them it is the most important – both then and now. They open chapter 2, titled *Complex systems and Mechanistic Explanation* with: 'Our aim is to develop a cognitive model of the dynamics of scientific theorizing that is grounded in actual scientific practice' (p17). This is the background against which they are interested in characterizing mechanistic explanation. Understanding this is vital to understanding all their work, including the better known characterization of mechanism.

For Bechtel & Richardson, drawing on work of Herbert Simon and William Wimsatt, mechanism discovery is a process whereby boundedly rational human beings use heuristics to make progress in otherwise impossible searches. Decomposition and localization are not central to their account of mechanism because they are conceptually central to mechanisms, or because they are essential assumptions for mechanistic explanation, or even because they are the only available approach, but *because they are the most widely used and effective heuristics*.

For Bechtel & Richardson, decomposition and localization are fallible strategies, useful to minds like ours faced with the particular problems of finding mechanisms: that the space of possible mechanisms is very large, so we require guidance; and, as above, that we need real constraints on possible answers to avoid generating too many too easily. Bechtel & Richardson write: 'In the cases we have considered there is no antecedently specifiable set of mechanisms, and hence no well-defined space to search for possible mechanisms. Indeed, we do not even know whether the space we are searching contains a solution at all. Decomposition and localization guide the search for an adequate model within that space; that is, they guide and constrain the construction of explanatory models. Another, distinct, limitation is critical here: the data vastly underspecifies the nature of the mechanism. Thus, the researcher must develop strategies for finding plausible models given the data. Decomposition and localization provide such a strategy.' (p236.)

It is easy to say that when faced with a difficult problem, we need heuristics. But Bechtel & Richardson deliver more than that. Understanding heuristics is just the first step, which allows us to modify our thinking and our methods in useful ways. First, heuristics are of course fallible. But heuristics are also problem specific, and may fail systematically, particularly out of context. Investigating such failures allows us both to construct better epistemic methods, and to evaluate existing methods better. Bechtel & Richardson turn naturally to Wimsatt's demand for robustness: a variety of fallible, but independent, methods to detect something might ultimately be more effective than a single highly reliable procedure which still has some probability of failure.

This is why the mutual constraints imposed by trying to map a functional to a structural decomposition are so important. Trying to map them shows you when you've got one or both of the decompositions wrong, and gives you evidence that you've finally got them both right. This is why it is such an important achievement to get a mechanistic explanation. It is also why, when you still have only a structural decomposition, or a functional decomposition, many reject the proposed explanation as too unconstrained to be convincing.

That decomposition and localization help even though they are fallible is more than enough to be interesting. But Bechtel & Richardson push their point even further. The heuristics of decomposition and localization help – they usefully guide search – even when the underlying assumption that the system is decomposable *fails*. They write: 'The failure of decomposition is often more enlightening than its success: it leads to the discovery of additional important influences on behavior.' (p24.) This is an extraordinary claim, which should have a big impact on the various debates about modularity, if it can be defended.

Bechtel & Richardson defend it effectively by examining empirical work, of which I can give only a taster here. One example they use is fermentation, showing how empirical work has moved around failed assumptions. The first assumption in mechanistic explanation is often to assume that a single component is largely responsible for the phenomenon, and this often fails. The next move is commonly to accept that more than one component is involved, but assume that components are relatively independent, and linearly organized. For fermentation, this too failed: 'We have noted that linearity is a natural simplifying assumption. Chains of reactions are more readily tracked and understood than ones involving cycles. However, like other constraints, this assumption can be abandoned when other considerations tell against it, or when it conflicts with the other constraints we have mentioned.' (p162.) Although many initial assumptions failed: 'Ultimately it was discovered that the fermentation system was minimally decomposable. Various reactions were mutually interdependent. If we experimentally disrupted these reactions, then even a normally fermentable substance would not ferment. Even more significantly, disrupting one step in the pathway would block others that preceded it.' (p170.) Bechtel & Richardson claim that mechanistic heuristics still helped guide search, even for a fairly complex system involving significant feedback, and empirical work eventually led to a reconceptualization of the phenomenon whereby minimal decomposability was recovered.

Bechtel's current work, defending what he calls *dynamical mechanistic explanation*, continues to engage with using mechanistic heuristics to explain complex systems. But the roots are already planted deeply here. Also, the heuristics of decomposition and localization are distinctively mechanistic. So although the concept of mechanism has changed over time, the current concept, involving the engagement of these heuristics, is still a concept of mechanism.

Conclusion

In conclusion, there is a great deal to be gained from *Discovering Complexity*, including many things I have not mentioned, such as the discussions of isolating and reconceptualising the phenomenon, and Bechtel & Richardson's 2010 reassessment of emergence and the need to recompose and situate mechanisms.

I will close by pointing to new questions raised by *Discovering Complexity*. Properly understanding Bechtel & Richardson's heuristics of mechanism discovery shows how much

weight rests on satisfying mutual, independent constraints, and on integration of evidence from multiple levels. These are the strategies that allow us to come to confident conclusions even using fallible methods. And the use of such strategies affects both our understanding of mechanistic explanation, and much empirical work that is actually taken to support the existence of particular mechanisms. But while Bechtel & Richardson point to many extremely useful historical examples, they do not give us a detailed philosophical examination of what these mean.

First, take 'independence'. The natural interpretation of independence is 'probabilistic independence', so that even fallible independent convergent evidence of a mechanism renders that mechanism very probable. But, as Bechtel & Richardson are well aware, so much work on mechanisms is highly integrated, with studies building on each other, and experimental tools requiring calibration, which in turn depends on previous experimental success, that clean probabilistic independence seems unlikely. So there is room to explore what 'independent' evidence amounts to, and what conclusions the kinds of evidence we actually get can support.

Secondly, how do we integrate evidence from multiple levels? Here, Bechtel & Richardson leave the notion of 'level' fairly intuitive, but seem to have in mind something like levels of organization. So we integrate evidence about the functioning of a whole with evidence about the functioning of its parts. They also allow that at least sometimes, the evidence we integrate comes from different disciplines, such as from cognitive science and from neuroscience. So we have here two ways in which evidence from multiple levels might count as in some sense independent, and so able to impose mutual constraints in a way that significantly increases our confidence in the result. However, there is clearly room for a great deal more work, of both philosophical and empirical significance.

Of course, that Bechtel & Richardson don't say a great deal about these questions is in a sense a criticism of the work, but it is friendly. It is only due to Bechtel & Richardson's work that we have uncovered these questions, understood their pivotal place in mechanistic explanation, and been given the outline that will allow us to bring the answers to bear on understanding mechanistic explanation.