EXPLANATION AND INTEGRATION IN MIND AND BRAIN SCIENCE

DAVID M. KAPLAN

Reviewed by Joe Dewhurst

Explanation and Integration in Mind and Brain Science
David M. Kaplan *(ed)*
Oxford: Oxford University Press, 2017, £45
ISBN 9780199685509

The debate over whether, and to what extent, psychological explanations can be considered to be 'autonomous' is an old one, dating back at least to the 1970s.[1] This new volume, edited by David Kaplan, promises not only to reinvigorate that debate, but also to refocus it, by shifting the emphasis away from an abstract dichotomy between autonomy and reduction, and towards analyses of specific explanatory practices in the mind and brain sciences. This allows for a more nuanced approach, where different levels of explanation are neither wholly autonomous nor entirely integrated, but rather exercise mutual constraints upon one another. Each of the contributors to the volume engages with this theme in some way, either by presenting a particular case study (Strevens, Kaplan, and Aizawa), considering specific ways in which different levels of explanation might constrain one another (Woodward, Egan, and Shagrir and Bechtel), or exploring the implications of a non-reductive approach to theoretical integration (Roth and

Cummins, Weiskopf, Murphy, Maley, and Piccinini). This is not to say that there is complete agreement amongst the authors. For example, Weiskopf defends the explanatory autonomy of cognitive models, Egan argues for the autonomy of function-theoretic models, and Aizawa considers examples of multiply realizable (and thus partially autonomous) kinds in the science of colour vision. Even in these latter cases though, the tone is generally conciliatory—every chapter of this volume makes an effort to move the debate forward rather than simply re-treading old ground, and it is perhaps most valuable for the questions it leaves unanswered, providing invitations for future debate and discussion.

Several chapters focus on topics relating to mechanistic explanation, which offers a distinctive perspective on autonomy and integration.[2] Egan considers the relationship between function-theoretic models and mechanistic explanations, arguing that the former cannot be fully integrated into the latter. Kaplan responds to a concern raised by Chirimuuta ([2014]) about 'canonical neural computations', which appear to be multiply realizable, and thus 'cannot be explained in mechanistic terms' (p. 165). Shagrir and Bechtel explore the role of phenomena in mechanistic explanation, focusing on quantifiable and contextually specified phenomena. Finally, Maley and Piccinini aim to develop a novel mechanistic account of teleological functions, based around the 'objective goals' of the target organism. These four chapters are grouped towards the end of the volume and taken together raise several outstanding issues that the mechanistic explanation framework will be faced with moving forward. I will return to these issues (and their associated chapters) shortly, but first I will briefly discuss one earlier chapter that makes an especially distinctive contribution and will be of particular interest to a non-specialist audience.

Dominic Murphy's chapter on 'Brains and Belief' presents a compelling analysis of the relationship between folk psychology and scientific psychology, focusing on three distinct 'perspectives' on this relationship: integration, autonomy, and elimination (pp. 121–5). Each perspective gives a different set of answers to the following three questions: 'does folk psychology make empirical commitments?', 'is folk psychology true (or alternatively, predictively and explanatorily powerful)?', and 'does folk psychology define the top level in an explanatory hierarchy?' (p. 125). Advocates of integration and autonomy agree on the second question, advocates of integration and elimination agree on the first, and advocates of autonomy and elimination agree on the third. Murphy goes on to consider each question in more detail, and the implications each perspective has for contemporary debates in philosophy of cognitive science (including those considered elsewhere in the volume). His own conclusion is that while integration is currently the dominant view, its future success is dependent on 'the extent to which key constructs of folk psychology can survive amendment in the light of neurological evidence, which threatens to dissolve our existing concepts and introduce new ones' (p. 141). The further these new concepts move away from our existing ones, the more integration begins to look like elimination. In this case, a defender of autonomy might be able to save folk psychology from elimination, but Murphy suggests that the real significance of eliminativism is likely to be ethical or political: 'If the new sciences of the mind reinterpret human beings too substantially, we will risk losing our grip on what matters to people' (p. 142). Even if one does not agree with Murphy's conclusions, they are certainly thought-provoking, and his analysis of the differing perspectives on folk psychology is clear and helpful. In an otherwise quite technical volume, this chapter is most likely to be of interest to the general reader.

As indicated above, mechanistic explanation is a major focus of the volume, especially in the four chapters discussed below. This is perhaps not surprising, given Kaplan's own research interests, but it also makes sense given the broader topic of the volume, namely, the tension between integration and autonomy in the mind and brain sciences. As Kaplan indicates in his introduction, previous attempts to resolve this tension have tended to emphasize a dichotomy between integration (or reduction) and autonomy, whereas at least some proponents of mechanistic explanation have tried to explore a middle ground between these two extremes.[3] The mechanistic approach to explanation is not without problems of its own, however, as the issues raised in these chapters make clear.

The chapters by Frances Egan and Kaplan himself both address a broadly related issue, which is the compatibility (or lack thereof) between mechanistic explanations and explanations that appeal to higher-level or more abstract generalizations. In 'Function-theoretic Explanation and the Search for Neural Mechanisms', Egan defends the existence of a distinctive form of 'function-theoretic' explanation that 'can be genuinely explanatory even absent an account of how the capacity is realized in neural hardware' (p. 145). Such explanations involve the characterization of a cognitive task in terms of 'an independently well-understood mathematical function' (p. 146), which, Egan argues, has its own kind of explanatory value, even when 'nothing in the function-theoretically characterized system corresponds to (states of) components of neural mechanisms' (p. 149). If Egan is correct, then mechanistic explanation cannot be the whole story.

In 'Neural Computation, Multiple Realizability, and the Prospects for Mechanistic Explanation', Kaplan responds to a related set of arguments against mechanistic explanation, which appeal to the explanatory success of 'canonical neural computations', that is, the abstract characterizations of computational/mathematical processes that are implemented by several different neural mechanisms (see, for example, Chirimuuta [2014]). He argues that mechanistic explanation can in fact accommodate certain kinds of multiple realizability, by allowing for abstractly characterized, multiply realizable phenomena that are nonetheless instantiated by concrete mechanisms. He focuses on the example of sound localization in birds and mammals, which in both cases involves similar (abstract) computations implemented in quite distinct (concrete) mechanisms. Crucially, Kaplan indicates how in each case the mechanisms involved are now relatively well understood, satisfying constraints on adequate mechanistic explanation without ruling out the possibility of multiple realization.

Both Egan and Kaplan consider whether, and to what extent, multiple realizability is compatible with the mechanists' '3M' constraint, previously introduced by Kaplan and Craver ([2011]), which requires that the elements of a mechanistic model correspond to 'identifiable components, activities, and organizational features of the target mechanism' ([2011], p. 611). Egan denies that this is necessary in the case of function-theoretic models (p. 160), whereas Kaplan identifies a case where two different models of a multiply realizable phenomenon can each satisfy the constraint (p. 180). These conclusions are not necessarily at odds with one another—Kaplan could accept that other (non-mechanistic) modes of explanation are viable, and Egan could grant that in some cases function-theoretic models might be able to satisfy the 3M constraint, even if she does not think that this is a necessary requirement on explanation.

In 'Marr's Computational Level and Delineating Phenomena', Oron Shagrir and William Bechtel consider the importance of environmental context for fixing the phenomenon to be investigated by a mechanistic explanation. (Egan also considers the role of environmental context, which she thinks can supplement an environment-neutral function-theoretic model, but she does not think it is essential to the function-theoretic mode of explanation.) Shagrir and Bechtel explore this in relation to Marr's computational level, which they argue should be understood as 'characterizing the phenomenon for which a mechanism is sought as explanation' (p. 201). Environmental context is important in determining the nature of the task that the mechanism is intended to solve; they describe this context as providing both methodological constraints on how we go about investigating the mechanism, and explanatory constraints on the kind of mechanism that would be suitable for solving the task. For example, the fact that we live in a world populated by mostly smooth objects places constraints on the kinds of mechanism suitable for visual processing, which would be different were we to live in a world populated mostly by jagged objects. Such constraints can play an important role in scientific discovery, as even prior to beginning to investigate the mechanism itself (in this case, the visual system) they can give us a sense of what that mechanism might be doing. Shagrir and Bechtel go on to describe how the initial characterization of a phenomenon might require updating in the course of developing a mechanistic explanation, as we uncover new constraints of which we were not previously aware. Here they make the interesting suggestion that Marr's approach might be supplemented with insights from the Gibsonian tradition, which could provide a richer analysis of the environmental (or in Gibsonian terms, 'ecological') context (see, for example, Gibson [1979]). This is certainly an aspect of the mechanistic approach to explanation that has been previously neglected, and Shagrir and Bechtel's chapter provides a solid foundation for future research in this area.

Finally, in 'A Unified Mechanistic Account of Teleological Functions for Psychology and Neuroscience', Corey Maley and Gualtiero Piccinini develop a novel account of teleological functions for neurocognitive mechanisms.[4] They claim that such an account is necessary, as 'neurocognitive mechanisms [...] appear to be *for* something: they appear to have *teleological functions*' (p. 237). Their account is based on what they call the 'objective goals' of an organism—survival and inclusive fitness—with a teleological function simply being any function that makes a stable contribution to an objective goal (pp. 243–4). This is similar in flavour to previous etiological accounts of function, but avoids making any appeal to evolutionary history, which they dismiss on the grounds of being both hard to investigate and causally impotent (pp. 238–9). While many will be sympathetic to these concerns and will agree that *some* account of function is required for mechanistic explanations, it is still too early to say how successful Maley and Piccinini's objective goal account will be. Perhaps most importantly, it might appear to some to be too close to etiological accounts, either risking a collapse into a version of those accounts or else suffering from some of the same shortcomings. Nonetheless, in coming up with a new proposal they have moved the debate forward, and the onus is now on the sceptics to explain *why* they are not convinced, and ideally to offer an alternative approach.

While these chapters do not conclusively resolve any of the three issues (multiple realizability, environmental context, and teleological functions), what they do manage to do is present the issues from an original and refreshing perspective, breathing new life into the now well-established mechanistic explanation framework. Each of the other chapters in the volume does something similar, offering engaging insights into the broader themes of explanation and integration, while also opening up new avenues for future investigation. I am confident that this collection will provide a valuable point of departure for many future debates in the philosophy of cognitive science, and I would highly recommend it to anyone looking to get involved in these debates.

Joe Dewhurst
Department of Philosophy
University of Edinburgh
joseph.e.dewhurst@gmail.com

References

Chirimuuta, M. [2014]: 'Minimal Models and Canonical Neural Computations: The Distinction of Computational Explanation in Neuroscience', *Synthese*, **191**, pp. 127–53.

Craver, C. F. [2007]: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Oxford University Press.

Craver, C. F. and Kaplan, D. [forthcoming]: 'Are More Details Better? On the Norms of Completeness for Mechanistic Explanations', *British Journal for the Philosophy of Science*.

Fodor, J. [1974]: 'Special Sciences (Or: The Disunity of Science as a Working Hypothesis)', *Synthese*, **28**, pp. 97–115.

Gibson, J. J. [1979]: *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin.

Kaplan, D. and Craver, C. F. [2011]: 'The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective', *Philosophy of Science*, **183**, pp. 1–35.

Piccinini, G. [2015]: *Physical Computation: A Mechanistic Account*, Oxford: Oxford University Press.

Notes

[1] With Fodor ([1974]) presenting the classic defence of the autonomy of psychology and the other special sciences.

[2] Consider, for example, Craver's ([2007]) proposal for a 'mosaic unity' in neuroscience, which would allow for integration between levels without rendering them fully reducible.

[3] See (Craver and Kaplan [forthcoming]) for a recent exploration of this middle ground in terms of the relationship between mechanistic explanation and abstract models.

[4] Although not entirely novel, as Piccinini presented a version of this proposal in his ([2015], Chapter 6).