2012

Mechanistic Explanations and Animal Model Simulations in Neuroscience

Nina Atanasova

Mechanistic explanations and animal model simulations in neuroscience

Nina Atanasova Atanasova

The 'new mechanists' assume that discovering neural mechanisms is a major aim of neuroscience where this constitutes a process of explaining. They strive to articulate the norms for good explanations. I argue that: the normative project of the ontic mechanistic explanation is unrealistic; the epistemological version of mechanistic explanation accounts for the practical epistemological constraints of neuroscience but fails normatively; hence a dilemma: either methodologically strong but unrealistic or epistemologically realistic but methodologically weak view of mechanistic explanation. I propose that the solution is in abandoning the idea that the study of mechanisms in neuroscience aims mainly at explaining. Model-building and simulating neural phenomena and mechanisms do not necessarily aim at providing explanations. Additionally, some of the cognitive goals attributed to explanation are actually served by simulation. Thus, the new mechanist approach needs to be revised so as to do better justice to the role of simulation in neuroscience.

1. Introduction

The so-called 'new mechanism' is perhaps the most promising among the contemporary approaches to the philosophy of neuroscience. Its proponents, by and large, assume that a major aim of contemporary neuroscience is discovering mechanisms underlying neural phenomena, where this is also a process of explaining. Thus, the main question that concerns them is: 'What constitutes a good mechanistic explanation?'. Two groundbreaking programmes are Craver's (2007) normative account of the ontic mechanistic explanation and Bechtel's (2008) epistemologically focused mechanistic reduction, according to which mechanistic explanations have representational rather than ontic status.

In this article, I analyse the two proposals and conclude that they are deficient in that they do not do justice to significant practices in neuroscience that do not aim at providing explanations. Rather, those practices aim at mimicking the functioning of the studied mechanisms. These practices seem to have cognitive value of their own apart from explanation. My purpose is to show that the new mechanism, as an account of explanation, needs to be supplemented by an account of the role of simulation in the

process of studying, modelling and manipulating mechanisms in neuroscience. Thus, I defend the following claims:

(1) The normative project of specifying the requirements for good explanations in neuroscience, as articulated by Craver (2007), is flawed at least because it does not fit the actual practice of neuroscience but arguably also because it prescribes unrealistic methods. It requires that good explanations be complete models of the real neural mechanisms that produce the studied phenomena. This requirement is what stands in the way of Craver's ontic approach because neuroscience hardly ever operates with complete models. Moreover, the actually used models are not generally considered as explanations.

(2) Even though Bechtel's account does justice to the partial character of the neuroscientific knowledge which results from the limited epistemic capacities of the neuroscience research techniques, it nevertheless fails to demonstrate how this partial knowledge plays a satisfactory explanatory role. Incorrect explanations, on this account, are as much acceptable explanations as they would be if they were correct. This is to say that the account is methodologically weak.

(3) Thus, one has to face what I call the 'Craver–Bechtel dilemma'. According to the dilemma, the available approaches to the study of the practice of neuroscience are either methodologically strong but unrealistic or epistemologically realistic but methodologically weak as views of mechanistic explanations. However, the dilemma could be overcome if one abandons the idea that the study of mechanisms in neuroscience is directed exclusively or even mainly towards producing (accurate) explanations and acknowledges that simulations of mechanisms generate valuable cognitive production of a kind significantly different from the traditional theoretical explanations. Thus, even though mechanistic explanations are not as strong as the ontic approach requires, their goals can still be accomplished, only through simulation.

(4) This result, in turn, suggests that some of the cognitive goals that were initially attributed to explanation should now be redirected towards simulation. The latter is a philosophically underappreciated practice in neuroscience which needs a better analysis from a philosophical perspective. Indeed, there is a philosophical discussion of some computational simulations in cognitive neuroscience but hardly any serious treatment of simulations with animal models for example. I make a case using an example from the research on depression and anxiety that employs simulations with animal models.

2. Ontic mechanistic explanation

The 'new mechanists' advance their view as an alternative to the classical positivist conception of scientific explanation, known as the deductive-nomological (D-N) or covering law model of explanation. They draw their inspiration from the systems approach and are mainly concerned with analysing the practice of the biological, neurobiological and psychological sciences. The main starting points of the approach are: (1) typically, explanations in the life sciences are provided in terms of identifying mechanisms responsible for the production of the explanandum phenomena rather than in terms of laws from which the latter can be derived; (2) thus, explanations are not arguments but exhibitions of the mechanisms underlying the explanandum phenomena and (3)

mechanistic explanations are supposed to account for multilevel complexities. While the aforementioned convictions are shared by a broader range of philosophers of science, there is a distinct portion of adherents to the new mechanist approach who advocate an ontic view of explanation.

The first articulation of the mechanistic explanation (in biology and neurobiology) as ontic explanation can be found in Machamer, Darden, and Craver (2000). They note that '[i]n many fields of science what is taken to be a satisfactory explanation requires providing a description of a mechanism' (Machamer et al. 2000, p. 1) where '[m]echanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions' (Machamer et al. 2000, p. 3) and explanations are descriptions of mechanisms in terms of the activities and entities involved in the production of the explanandum phenomena. In their view, 'To give a description of a mechanism for a given phenomenon is to explain that phenomenon, i.e. to explain how it was produced' (Machamer et al. 2000, p. 3). The distinctive feature of the ontic explanation is that in order for it to count as an explanation, it needs to be revealing of the real mechanism responsible for the explanandum phenomenon.

The view is further developed in Craver and Darden (2001) where they elaborate on the distinction (already present in Machamer et al. 2000) between mechanism schemata and mechanism sketches.

> Mechanism schemata are abstract descriptions of mechanisms that can be instantiated to yield descriptions of particular mechanisms... Schemata are thus complete in the sense that they can be filled in without gaps in the productive continuity of the mechanism... Mechanism sketches, in contrast to mechanism schemata, are abstract descriptions of mechanisms that cannot yet be filled in. Mechanism sketches have black boxes – they leave gaps in the productive continuity of the mechanism... Such black boxes in mechanism sketches are useful in providing guidance about where further elaboration is needed. (Craver and Darden 2001, p. 120)

The distinction is important because it provides a basis for formulating criteria for good explanations, which as explicated by Craver, are by definition complete explanations that contain 'all and only' the relevant components of a mechanism (Craver 2007, p. 140). In addition, the distinction (introduced in Machamer et al. 2000, p. 21) between how-possibly, how-plausibly and how-actually descriptions of mechanisms is developed in Craver (2006, 2007) and it also plays an important role in his formulation of the criteria for good mechanistic explanations conceived as models. In Craver's view, explanatory models range over how-possibly, how-plausibly and how-actually models. Not all models are explanatory, according to Craver, and the explanatory models come in varieties as well (Craver 2006). Here are the three important varieties:

> *How-possibly models* have explanatory purport, but they are only loosely constrained conjectures about the sort of mechanism that might suffice to produce the *explanandum phenomenon*. They describe how a set of parts and activities might be organized together such that they exhibit the *explanandum phenomenon*. One might have no idea if the conjectured parts exist or, if they do, whether they are capable of engaging in the activities attributed to them in the model... How-possibly models are often heuristically useful in constructing and exploring the space of possible mechanisms, but they are not adequate explanations. *How-actually models*, in contrast, describe real components, activities and organizational features of the mechanism that in fact produces the phenomenon. They show how a mechanism works, not merely how it might work. Between these extremes is a range of *how-plausibly* models that

> are more or less consistent with the known constraints on the components, their activities and their organization. (Craver 2007, pp. 112–113)

It is clear that nothing less than a complete how-actually model would do for a good mechanistic explanation. This requirement however is so strong that it is doubtful that it is ever met in neuroscience where research is, as a rule, carried out on partial models of artificial cell cultures or highly idealised and speculative models and simulations. It is more likely that how-plausibly models are the best that neuroscience ever achieves.

Additionally, there is an important peculiarity of the ontic explanations which is implicit in the earlier articulations of the ontic view of mechanistic explanation but becomes explicit in Craver's exposition. It stems from the fact that the term 'explanation' may refer to the actual mechanisms underlying the real phenomena but it can also refer to the descriptions of those mechanisms. Craver calls the former type 'objective explanations'. Importantly:

> Objective explanations are not texts; they are full-bodied things. They are facts, not representations. They are the kinds of things that are discovered and described. There is no question of objective explanations being "right" or "wrong," or "good" or "bad." They just are. (Craver 2007, p. 27)

The second type of explanation Craver calls 'explanatory texts'. Those are the descriptions of the mechanisms. They are representational and they can be good explanations only in virtue of correctly describing the actual mechanisms. As such, they could be complete or incomplete. However, the explanatory texts are good explanations only if they are *ideally* complete (i.e. if they are describing all the relevant aspects of the mechanisms) how-actually descriptions of the real mechanisms producing the explanandum phenomena. Craver (2006) provides a detailed account of representational explanations conceived as models. He distinguishes them from non-explanatory models and specifies the requirements for explanatory models, which he terms 'mechanistic models' (Craver 2006, p. 367). Among other things, it is required that mechanistic models 'account fully for the explanandum phenomena' (Craver 2006, p. 368) and describe the real mechanisms that produce them.

> In order to explain a phenomenon, it is insufficient merely to characterize the phenomenon and to describe the behavior of some underlying mechanism. It is required, in addition, that the components described in the model should correspond to components in the mechanism in T [some feature of a system targeted for being modelled]... If one is trying to explain the phenomenon, however, it will not do merely to describe some mechanisms that would produce the phenomenon. One wants the model, in addition, to show how T produces the phenomenon. (Craver 2006, p. 361)

In fact, the ambiguity of the term 'explanation', as it figures in the ontic approach, is not just double but triple. There are three things that qualify as explanations on Craver's account. In addition to (1) mechanisms or objective explanations and (2) their descriptions or explanatory texts, there are (3) representational non-linguistic, often physical, models of mechanisms that share some (but not all) features with objective explanations and others with explanatory texts.

The addition of non-linguistic representations of mechanisms into the picture is necessary because much of science, and especially neuroscience, is preoccupied with model-building or experimenting on physical models of the explanandum phenomena. On Craver's account, this would mean that an explanation–construction in the case of model-building would require literal physical putting together of entities that are already

established as building blocks of mechanisms in the so-called 'ontic store', a notion discussed in Craver and Darden (2001) and Craver (2007). A good explanation in this case would be a complete how-actually model that would contain all and only the relevant parts of the objective mechanism. Eventually, this model would have to be an actual physical mechanism which would be an exact replica of the real mechanism out there in the world. It would be a purified extraction, so to speak, from the real phenomenon such that it would contain only its mechanism's productive entities and interactions.

Craver might in fact be inclined to accept such an extension of his account. For example, he admits that, on an externalist view of explanation, one could consider the computational tools and databases, which scientists use to capture very complex phenomena, as extended explanations. However, he does not trace the implications of this potential step (Craver 2007, p. 33, note 7). I will discuss some of these implications in the following section and will show how they may pose serious problems for Craver's account of mechanistic explanation.

3. Problems of the ontic view

It is clear that Craver has to accept the third type of explanation because otherwise he would be advancing an inconsistent account of explanation. Since physical models share relevant features with both objective and textual explanations, they have to be given status of explanations as well. However, they cannot be subsumed under either of the two established categories. A physical model of the mechanism under study that contains all and only the relevant parts of the real mechanism will be the same mechanism as the real objective mechanism. Nevertheless, it is constructed and because it is constructed with the intent to *exhibit* the real mechanism in a clearer fashion, it is also a representation.

However, granting to physical models the status of explanations would pose another problem for Craver's view. If he accepts the third type of explanation, his account may eventually lose its descriptive adequacy. Scientists do not call such constructed physical mechanisms 'explanations'. They are more likely to call them 'simulations'. Now, if Craver insisted on the neat categorisation that his account of the role of mechanistic models suggests, he would lose sight on what model-building's actual purpose in the practice of neuroscience is. Craver's account would thus be normatively, that is methodologically strong, but descriptively inadequate. This result, coupled with the ambitious expectation that neuroscience aims at producing complete models of the real mechanisms would render the account highly unrealistic.

4. Mechanistic explanations as representations

Rejecting the strong metaphysical commitments of the ontic mechanistic explanations, Bechtel's epistemologically framed version of the mechanistic approach sidesteps the problem of unrealistic expectations that is inherent in Craver's ontic view. Bechtel's account does justice to the epistemic constrains that restrict the possible knowledge-generating procedures that neuroscientists can in fact perform in their practice. Bechtel stresses the importance of the 'mental labour' that neuroscientists employ in representing and explaining mechanisms (Bechtel 2008). He admits that it is not the real mechanisms in the world that explain the phenomena of interest. It is the representations of those mechanisms that play the explanatory role. Explanations in neuroscience are only constructions that are different from the real mechanisms they purport to explain.

They are constructed with the goal to be accurate but they are not infallible. Still an explanation is as much an explanation when it is inaccurate as it would be if it were accurate.

Bechtel proposes a revision of Craver's ontic view of mechanistic explanation which renders the latter a representation. Thus, 'a good explanation would be one that accurately represents the mechanism that actually produces the phenomenon' (Bechtel 2008, p. 19, note 6). However, given the epistemic constraints, one can never be certain whether or not one has represented the actual mechanisms producing the phenomena of interest. Bechtel's account is then methodologically weak. It does not specify clear criteria for judging the value of a given mechanistic explanation.

5. The Craver–Bechtel dilemma

Thus, a new mechanist interested in specifying the norms of good explanations in the actual practice of neuroscience faces the following dilemma:

(1) accept either Craver's or Bechtel's view;
(2) if Craver's view, then methodologically (normatively) strong but epistemologically unrealistic view of mechanistic explanation;
(3) if Bechtel's view, then epistemologically realistic but methodologically weak view of mechanistic explanation;
(4) therefore, either methodologically strong but epistemologically unrealistic or epistemologically realistic but methodologically weak view of mechanistic explanation (1, 2 and 3).

But remember that the project of specifying the norms of good mechanistic explanations was meant to be, like the D-N model of explanation, methodologically, that is normatively, strong. However, it was also meant to be, unlike the D-N model, epistemologically realistic (in that it had to be descriptively adequate and to propose a methodology for actually doing neuroscience). It was supposed to articulate the actual norms that are and should be followed by the neuroscientists in their practice. This means that the conclusion of the Craver–Bechtel dilemma is unacceptable.

Nevertheless, the new mechanist project is not yet to be abandoned. There may actually be a way to overcome the seemingly inextricable dilemma. I suggest that we look for the solution in the implications of a revised version of Bechtel's view.

According to Bechtel, an adequate account of scientific explanation should approach its conceptualising from an epistemological perspective. Mechanistic explanations are not an exception. They are the products of an elaborate cognitive activity on the part of the scientists, who construct them in search of understanding of the real world. Nevertheless, the explanations are only constructions that are different from the real mechanisms they purport to explain.

Bechtel points out that in addition to propositions, scientists often employ diagrams to represent mechanisms. This kind of representation is fundamentally different from the propositional representations that characterised the D-N model. The representations employed by the scientists when dealing with mechanistic explanations are more often visual representations. Bechtel argues that this is so for a good reason. When explaining a multidimensional dynamic mechanism, diagrams and even animations are much more suitable than linguistic descriptions. An important consequence of this observation is that

logic cannot account for the actual processes of reasoning in which scientists engage. This is due to the fact that logical reasoning presupposes propositional reasoning. Since, in many cases, it is not propositions that scientists use to come to their conclusions, but visualising techniques, the way they reach their conclusions cannot be logic. Simulations seem to play an important role that is fundamentally different from the role linguistic representations can play.

> Mechanistic explanation can, accordingly, be understood as an epistemic activity of scientists without restricting it to the sort of epistemic activity envisaged in the D-N approach to explanation. In addition to linguistic representation and inferences that can be characterized in logical terms, they may involve visual presentations and inferences that take the form of simulating the activity of the mechanism. (Bechtel 2008, p. 21)

In what follows, I take Bechtel's suggestion for a distinct cognitive role of simulation one step further and stress the fact that (at least) some important non-linguistic models in neuroscience do not purport to explain. Rather, they provide valuable insights through simulation. Their purpose is to exhibit workings of mechanisms without thereby explaining. This, however, does not make them any less valuable than explanations, as Craver implies. They simply serve a different cognitive purpose.

6. Representations without explanations: a case of simulations with animal models

Abandoning Bechtel's assumption that the visual representations of mechanisms in neuroscience are explanatory is the way to overcome the dilemma. I propose that it is a fruitful revision of Bechtel's account of the role of representations of mechanisms in neuroscience to preserve the term 'explanation' only for linguistic, propositional representations of mechanisms. They can still be expected to comply with the standard logical constraints. For example, linguistic explanations would still be expected to be coherent. Also, different descriptions of the same mechanisms would still be expected to be consistent with one another.

One might worry that this would bring us back to where we started, namely to the inadequate D-N model. This is not necessarily so, however. Even if we assume that explanations are arguments, we do not have to thereby assume that they are a syllogistic type of argument such as the one prescribed by the D-N model. The mechanistic explanations could be framed as causal arguments which would not require subsuming the explanandum phenomena under generalisations (whether universal or statistical) and initial conditions. This would well capture the causal features of the mechanistic explanations. It is not impossible to incorporate the other key feature of the mechanistic explanation, the constitution relation, in similar arguments as well.

Still, there would be numerous situations in which linguistic representations would not be capable of capturing the relevant aspects of the studied mechanisms. In those cases, visualising and model-building would be more appropriate tools for capturing the phenomena and the mechanisms that produce them. Simulation and not logic, or explanation thereby, would be the relevant activity employed in those cases.

There is an interesting story yet to be told about the models and simulations that the neuroscientists actually use. For various reasons, most of their experimental research is performed on models of the systems they study. Animal models are one such broadly utilised type in both medical and cognitive neuroscience. The major assumption employed in this kind of research is that given the common evolutional ancestry of the species

involved, they probably share common neurological mechanisms. Thus, the more basic the studied neural system, the more evolutionary distant the animal used as a model could be. Rodents (usually mice and rats), for example, are commonly used as models for depression. They are considered adequate models because they share behavioural similarities with humans and also, being mammals, they share homologous brain structures with humans.

Rodent animal models of depression have been largely employed in drug-screening assays. The forced swim test (not to be confused with Morris water maze test, which is largely employed in the study of spatial memory and learning) is commonly utilised for this purpose. In this test, a rodent is placed in a cylinder filled with water from which the rodent cannot escape without help. In the end of a session of useless attempts to escape, the rodent stops trying. After a prolonged subjection to such conditions, it would not try to escape from the water even if it were presented with a platform through which it could escape on its own. This behaviour is taken to mimic the so-called 'learned helplessness' symptom of depression. When treated with antidepressants known to be effective, the rodent does not exhibit the learned helplessness symptom.

Thus, new pharmaceuticals can be tested for effectiveness in the described set-up. This function of the rodent animal model is straightforward and it obviously does not strive to provide explanations. Nevertheless, it does not seem to present any problem for Craver's account either. It could well fit under the category of a phenomenal model, discussed in Craver (2006). This type of model is only supposed to mimic the overall phenomenon under study. It however does not require that the mechanism that produces the modelled phenomenon should be the actual mechanism that produces the phenomenon under study. Phenomenal models can be how-possibly models but may as well have absolutely no explanatory purport whatsoever.

There is, however, another function of animal models in neuroscience research. Animal models are often used to simulate human behaviours in search of understanding the workings of the human brain. Willner (1991a), for example, argues that animal models have a role, distinct from just drug-testing, in the study of depression. He points out that animal models are increasingly 'used as simulations for investigating the psychobiology of depression, and a minority of models have been developed explicitly for this purpose' (Willner 1991a, p. 131). Willner talks about the importance of this use of animal models in providing clinical insights and developing behavioural neuroscience.

More recently, Sufka, Weldon, and Allen (2009), Warnick, Huang, Acevedo, and Sufka (2009) and Hymel, Sameto, Kim, and Sufka (2010) among others argue that even domestic fowl chicks can be used as sufficiently reliable models of anxiety and depression. They argue that the avian model is even better than the rodent model because it accounts for the growing amount of data suggesting that anxiety and depression are in fact just temporally different stages of one and the same disorder.

What should strike one as highly problematic here is that the avian brain is structurally very dissimilar from the mammalian brain. At the very least, it does not have neocortex. How then could it be used as a model of the human brain? It is clear that one cannot claim the strict one-to-one correspondence between all and only the relevant parts of the studied mechanism and its corresponding model, as required by Craver's account. Fortunately, there are structures in the avian and mammalian brains that seem to share sufficient functional similarities. The chicks also exhibit distinct behavioural reactions to the experimental set-up that mimic relevant human reactions to a significant degree.

The model described in Sufka et al. (2009), Warnick et al. (2009) and Hymel et al. (2010) involves simulation of anxiety and depression symptoms induced through social separation. About 4–6-day-old socially raised chicks are taken away from their fellow-chicks. The isolated chicks exhibit high-distress vocalisation rates for the first 5 min of the test. This is considered as a sign of panic-like anxiety and is also an attempt to establish social contact with conspecifics. Within the next 15–20 min, the rate of the vocalisations drops in half and the chicks enter the stage of depression. The rate of distress vocalisations remains stable for the remainder of the experiment, which lasts 2 h in total. Since the vocalisations aim at establishing social contact, the drop in their rate is taken to model the learned helplessness symptom of depression. The chicks give up on trying establishing social contact because they have been unsuccessful in establishing it for a while. When treated with known anxiolytics (drugs effective in the treatment of anxiety), the chicks do not exhibit the anxiety symptoms, that is, their distress vocalisations are fewer in the first phase of the experiment compared to the untreated chicks. When treated with known antidepressants, the chicks do not exhibit the dramatic drop in distress vocalisations that characterised the stage of depression in the initial set-up.

What has been established about this model, so far, is what is known as *face validity* and *predictive validity* (categorisation first introduced by Willner 1986, 1991b). *Face validity* refers to the overall similarity between the phenomenon (distress vocalisations resulting from social separation and the drop of their rate) exhibited by the model and the explanandum phenomenon (anxious and depressed behaviour in humans). *Predictive validity* refers to the predictable effects that known pharmaceuticals produce in the animal model. Still, Craver could claim that so far it has only been established that the particular animal model is a good phenomenal model. However, there is another criterion that needs to be met for the model to be considered a good mechanistic model. Additionally, it needs to exhibit the proper mechanism that produces the studied phenomenon. So far, it has only been shown to be a phenomenal model, which is not explanatory even on Craver's account.

Indeed, there is such a consideration that neuroscientists working with animal models have as well. They also consider a further requirement to be met by their models. The requirement in question is what is known as the model's *construct validity* (introduced also by Willner 1986, 1991b). *Construct validity* is supposed to control for the actual mechanisms producing the phenomena of interest. In the case of the domestic fowl chick model of the anxiety–depression continuum, this is the measurement of the blood concentration of biomarkers that are known to be elevated in stressful situations and depressive conditions, respectively. Corticosterone, which is established as a biomarker of stress responses was increased in the first 15 min of the test. On the other hand, cytokine interleukin-6, which is implicated in the anxiety–depression continuum hypothesis, was found to be increased at the final stage of the test, thus confirming the model's construct validity (Sufka et al. 2009, pp. 531–532; Warnick et al. 2009, p. 153; Hymel et al. 2010, pp. 90–91).

This seems sufficient to provide both Craver and Bechtel with reasons to consider the model as an explanation of the human anxiety–depression continuum. It mimics the studied phenomenon, it makes reliable predictions and the mechanism implicated in the production of the studied phenomenon is implicated in the production of the modelled phenomenon as well. However, such a conclusion should strike one as implausible. Neuroscientists call this type of experiment 'simulations' and they testify to gaining insights and understanding from them. Certainly, one could claim that those are synonyms

of 'explanation' in the scientific vocabulary, but we have good reasons to suppose that explanations are just one form of understanding among others. Here is an example of the evaluation of the significance of the simulations with animal models:

> Animal models of behavior have helped us to understand the human syndromes of anxiety and depression by guiding our understanding of the syndromes' multifaceted symptoms, by providing insight into the etiological factors which initiated the shift to more target specific drugs, and through facilitating research of the physiological responses to novel therapeutic drug treatments. (Hymel et al. 2010, p. 86)

In conclusion they also state:

> ...animal models are an essential element in experimental research. In addition to understanding anxiety and depression separately, the chick anxiety-depression continuum model lends itself as a tool to examine anxiety and depression as being on a single multifaceted continuum. Further, the chick model is an important avenue for exploring the underlying neural substrates and treatment options associated with anxiety and depression. (Hymel et al. 2010, p. 104)

It is clear that exploring neural substrates and treatment options have a value different from explanation. Craver himself would agree that this is the case. He admits that neuroscience 'is driven by two goals' – explanation and control of the brain and the central nervous system (Craver 2007, p. 1). Also, he notes that:

> These two goals of neuroscience are complementary. Explaining the brain is one way to figure out how to manipulate it, and manipulating the brain is one way to discover and test explanations. (Craver 2007, p. 2)

This article aimed at doing justice to the second of the two goals of neuroscience. Aiming at specifying the norms of explanation, the new mechanists have lost sight of the complementary goal of neuroscience – manipulating the brain and the nervous system as a separate activity. It seems clear now that the two activities associated with the two goals of neuroscience need to be studied in parallel in order to capture the actual practice of neuroscience.

7. Conclusion

In this article, I examined two of the currently leading approaches to explanation in neuroscience. I pointed out that the implications of both Craver's ontic view of mechanistic explanation and Bechtel's epistemological view of mechanistic explanation are in need of revision in order to serve the two main purposes of the new mechanistic approach: to specify the norms for doing neuroscience and to be true to the actual practice of neuroscience. At present, they form an unacceptable dilemma for the new mechanists, a dilemma that requires abandoning one of the two goals mechanists share. I showed that some of the goals the new mechanists attribute to explanation are actually achieved through simulation. Among those goals is constructing replicas of the mechanisms that underlie the studied phenomena in the form of non-linguistic representations. I supplemented this criticism with a concise case study of the role of an animal model as a simulation of the anxiety–depression continuum. In the analysis of the case study, it becomes clear that some of the normative criteria that Craver claims to be implicitly directed towards explanations are in fact quite explicit and reflected on, only they are norms for good simulations.

Acknowledgements

I was inspired for writing this article in Robert Richardson's seminar on Philosophy and Cognitive Science, University of Cincinnati, Winter Quarter 2010. A version of this article was first presented as a poster at the Epistemology of Modeling and Simulation conference, 1–3 April 2011, Pittsburgh, PA, USA. Later versions were presented at ISHPSSB, 10–15 July 2011, Salt Lake City, UT, USA and PhiloSTEM-2, 11–12 November 2011, Fort Wane, IN, USA. I benefitted greatly from the discussions at those meetings. I especially thank Lilia Gurova who generously provided me with substantial comments at all stages of drafting the article. I also thank John Bickle for thorough comments on a mature draft.

References

Bechtel, W. (2008), *Mental Mechanisms: Philosophical Perspectives on Cognitive Neuroscience*, New York: Taylor & Francis Group.

Craver, C. (2006), 'When Mechanistic Models Explain', *Synthese*, 153, 355–376.

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

Craver, C., and Darden, L. (2001), 'Discovering Mechanisms in Neurobiology: The Case of Spatial Memory', in *Theory and Method in Neuroscience*, eds. P. Machamer, R. Grush and P. McLaughlin, Pittsburgh, PA: University of Pittsburgh Press, pp. 112–137.

Hymel, K.A., Sameto, A.L., Kim, E.H., and Sufka, K.J. (2010), 'Development and Validation of the Chick Anxiety-Depression Continuum Model', in *Translational Neuroscience in Animal Research*, eds. J.E. Warnick and A.V. Kalueff, New York: Nova Science Publishers, pp. 83–110.

Machamer, P., Darden, L., and Craver, C. (2000), 'Thinking About Mechanisms', *Philosophy of Science*, 67, 1–25.

Sufka, K.J., Weldon, M., and Allen, C. (2009), 'The Case for Animal Emotions: Modeling Neuropsychiatric Disorders', in *The Oxford Handbook of Philosophy and Neuroscience*, ed. J. Bickle, Oxford and New York: Oxford University Press, pp. 522–536.

Warnick, J.E., Huang, C.J., Acevedo, E.O., and Sufka, K.J. (2009), 'Modeling the Anxiety-depression Continuum in Chicks', *Journal of Psychopharmacology*, 23, 143–156.

Willner, P. (1986), 'Validation Criteria for Animal Models of Human Mental Disorders: Learned Helplessness as a Paradigm Case', *Progress in Neuropsychopharmacology and Biological Psychiatry*, 10, 677–690.

Willner, P. (1991a), 'Animal Models as Simulations of Depression', *Trends in Pharmacological Sciences*, 12, 131–136.

Willner, P. (1991b), 'Behavioural Models in Psychopharmacology', in *Behavioural Models in Psychopharmacology: Theoretical, Industrial and Clinical Perspectives*, ed. P. Willner, Cambridge, New York, and Melbourne: Cambridge University Press, pp. 3–18.