How radical is too radical? Chemero's epistemological arguments against mental representation

Whit Schonbein
schonbeinw@cofc.edu

College of Charleston

Presented at the 2011 annual meeting of the Southern Society for Philosophy and Psychology, New Orleans, March 9-11, 2011.

ABSTRACT

In his recent book, *Radical Embodied Cognitive Science* (2009), Anthony Chemero offers two epistemological arguments against the explanatory utility of mental representation in cognitive science. These arguments are based on a contrast between 'dynamical explanations' (DEs), which utilize dynamical systems theory and do not contain representations as part of their theoretical apparatus, and 'representational explanations' (REs), which posit contentful states. The 'Predictive Argument' claims that that since DEs can predict the behavior of the target system, REs are superfluous. The 'Dependency Argument' claims that REs depend on DEs, and hence do not add to our understanding. In this paper I argue that both of these arguments fail: REs *do* add to our understanding, and REs are *not* dependent on DEs. Instead, I suggest DEs be viewed as working in tandem with REs in such a way as to promote new approaches to understanding representation.

1. Introduction

In his recent book, *Radical Embodied Cognitive Science* (2009), Anthony Chemero adopts a novel approach towards rejecting the explanatory utility of representations in cognitive science. Whereas previous anti-representationalist arguments have focused primarily on the metaphysical issue of whether cognitive systems use representations (Van Gelder, 1995; Bechtel, 1998), Chemero proposes to remain metaphysically agnostic, offering instead *epistemological* arguments against representation.

Chemero's arguments begin with a contrast between dynamical explanations (DEs) and representational explanations (REs). DEs (i) use the mathematical tools of dynamical systems theory to analyze the behavior of a target system in terms of coupled sets of equations tracking state changes over time, and (ii) do not include representations as part of their theoretical apparatus (Van Gelder, 1995; Eliasmith, 2003). In contrast, REs explain by showing how contentful states of a system interact to produce behavior.

Given this contrast, Chemero argues that DEs render REs superfluous, for two reasons: First, the 'Predictive Argument' claims that, since DEs can predict the behavior of the target system, there is nothing left for REs to do. Second, the 'Dependency Argument' claims that REs depend on DEs, and hence do not add to our understanding of the target system.

In this paper I argue that both of these arguments fail: REs *do* add to our understanding, and REs are *not* dependent (in any robust sense) on DEs. Instead, I suggest DEs be viewed as working in tandem with REs in such a way as to promote new approaches to understanding representation.

2. The Received View of Representation

Representational explanations posit states that carry content. Chemero adopts Millikan's (1989, 1996) analysis as "a traditional theory of representation" (Chemero, 2009, p. 50). However, his argument rests primarily on the *teleological* aspect of Millikan's account, which is shared by related theories.

Roughly speaking, the teleological aspect is this: State *C* of a system is a representation when it has the function of indicating feature *F* of the environment. For Dretske (1988), *C* first has *natural meaning* arising from reliable causal relationships with *F*, and then, through ontogenetic or phylogenetic adaptation, subsequent processes *M* come to *use C* for its capacity to indicate *F* (p. 62, p. 84). For Millikan, the 'consumer' *M* determines the content of *C* so that natural meaning "drops out" of the equation (Millikan, 1989, p. 246). In either case, *C* is a representation by virtue of the role it plays mediating between *F* and *M*. Call this the 'received view' of representation.[1]

These modest conditions on being a representation explain why Chemero opts for an epistemological approach. Because they are not difficult to satisfy, *granting* the received view comes close to making antirepresentationalism false by definition (Chemero, 2009, p. 67). On the other hand, if we try to strengthen the account (specifically, by making 'decoupling' a necessary condition on representation), the resulting conditions are *too* strict in the sense they are hard to defend (Chemero, 2009, chapter 3). Thus Chemero *grants* that DEs can be given representational glosses, and the issue turns on whether these REs add anything to our understanding of the system.

On my reading, Chemero offers two arguments for why REs do *not* add significantly to the understanding granted by DEs. These arguments arise in the context of a discussion contrasting dynamical and representational explanations of a

---

[1] Note that 'action-oriented' or 'pushmi-pullyu' representations can be accommodated by the received view. According to Millikan, such representations are special cases where the representation is 'fused' with action: "Its function is to mediate the production of a certain kind of behavior such that it varies as a direct function of a certain variation in the environment, thus directly translating the shape of the environment into the shape of a certain kind of conforming action" (1996, p. 151). These seem to be clear cases where consumers of intermediate states use them for their capacities to carry information about the world, and therefore cohere with the received view.

robot (Husbands, Harvey, & Cliff, 1995; henceforth HHC), so it will be useful to have a brief sketch of this work.

HHC construct a simulated robot and circular arena, and the robot is given the task of navigating to the center of this arena. The robot has tactile and visual sensors that detect contact with walls and lighting, and two wheels for movement. The controller of the robot is an artificial neural network (ANN) with a fixed number of input (sensor) and output (motors) nodes, but the number of hidden nodes and the connections can vary. Connections are both excitatory and inhibitory, and each node is noisy. The ANN controller is constructed indirectly using a genetic algorithm. Having evolved a controller that successfully solves the task, HHC proceed to analyze the controller using the tools of dynamical systems theory, beginning with the dynamics of single nodes, and culminating in sets of equations – a DE – that characterize the behavior of the robot in all possible sensory conditions. The issue, then, is whether a RE makes any additional contribution to this DE.

Chemero gives two reasons in support of a negative answer:

> There are reasons to prefer the dynamical account to the representational one described above. First, the representational story depends on the dynamical story about the control system [of the robot], but not vice versa. … [I.e.,] the representational description is dependent upon the dynamical one. [Second,] the representational description of the system does not add much to our understanding of the system. Once we have the full dynamical story, we can predict the behavior of the robot in its environment completely, and we can do so without making reference to the representational content of any states of its control system. … [I.e.,] the representational gloss does not predict anything about the system's behavior that could not be predicted by the dynamical system alone. (Chemero, 2009, p. 77)

The first argument – the 'Dependency Argument' (DA) – rejects representational explanation on the grounds that it is dependent on prior dynamical explanation. The second argument – the 'Predictive Argument' (PA) – rejects representational explanation on the grounds that it does not predict any behavior beyond that predicted by dynamical explanation. In the next two sections, I examine each of these arguments, in reverse order.

3. The Predictive Argument

The predictive argument can be summarized:

> (PA) DEs fully predict the behavior of a target system, and if a DE is fully predictive, then an RE is superfluous. Therefore, the RE is superfluous.[2]

---

[2] The PA comes up several times in chapter 4: "Why should we bother with representational explanations when we have precise, perfectly general, counterfactual-supporting mathematical

What is immediately striking about the PA is that it focuses exclusively on *prediction* as our explanatory goal. However, it is widely held that teleological explanations (of which REs are a subset) do more than predict.[3] For example, Canfield (1964) notes that they explain *why* a system has a given component, and they do so in a way that contrasts with covering law explanations:

> [T]he question 'Why do animals have livers?' might be interpreted as a request for a covering law explanation of why livers are present in animals. On the other hand, it might be interpreted as asking what the liver does, or what it is good for. When we give a teleological explanation in answer to the question, we have interpreted it in the second and not the first way. (Canfield, 1964, p. 295)

Dretske (1988) also makes this point while discussing the conditions under which an indicator *C* of feature *F* becomes a representation, noting that *C* also becomes explanatorily relevant to understanding *why* it causally interacts with consumer *M*:

> [The] causal relationship between *C* [the indicator] and *M* [the consumer] will have to be explained by the fact that *C* indicates, or has the function of indicating, how things stand elsewhere in the world. It will not be enough merely to have a *C* that indicates *F* cause *M*. We want the fact that it indicates *F* to be an explanatorily relevant fact about *C* – the fact about *C* that explains, or helps explain, *why* it causes *M*. … An indicator element … becomes a representation by having part of *what* it indicates … promoted to an explanatorily relevant fact about itself. (Dretske, 1988, p. 84).

That is, in the same way that identifying the *function* of a component helps explain why that component is present in the system (Canfield), identifying the function of a component as that of *representing F* helps explain why that component is present (Dretske). This sort of information is *not* given by a DE, which, by Chemero's own admission is non-teleological.

Bechtel (1998) applies the same observation to Van Gelder's (1995) claim that representation-talk lacks explanatory utility when describing the behavior of the Watt governor.[4] Bechtel notes that, "[t]he fact that the angle of the spindle arms

---

ones?" (Chemero, 2009, pp. 72-73). And, "If one has the complete dynamical story, what is left to be explained?" (Chemero, 2009, p. 73)

[3] This is not to say that representational explanations cannot predict behavior – propositional attitude ascriptions are an obvious example.

[4] Because this example has been discussed at length in the literature, I assume the reader is familiar with the construction of the Watt governor. If not, the task of a governor is to control the speed of a steam engine. To keep the speed constant, a valve is used to release steam when the engine is going too fast, and closed when the engine needs to speed up. The engine is connected to a flywheel, which spins a set of weighted arms, which will rise due to centrifugal force as the flywheel increases in speed. The arms are connected to the valve so that the higher they rise, the more the valve is opened, and the lower they fall, the more the valve is closed. In this way the mechanism – the governor – automatically corrects the speed of the engine.

represents the speed of the flywheel becomes more clear when we consider *why* it was inserted into the mechanism to begin with." (p. 303, italics added). Bechtel's point is that the spindle was constructed the way it was *because* that design carries accurate information about the speed of the flywheel. Suppose, for example, the design was modified in some way, e.g., by changing the material composition of the arms – in that case the overall causal dynamics of the system would be of the same general sort. But the mechanism would not longer meet the task demands because the modified spindle no longer accurately represents the flywheel speed. Consequently, without acknowledging the representational function of the arms we are left with only a 'causal skeleton' describing the interactions of the components, but not the reasons behind them.

So, there appears to be a consensus that teleological explanations in general, and representational explanation in particular, answer a 'why' question about the presence of components in a system, one that is not answered by dynamical explanations.[5]

4. The Dependency Argument

A first stab at the DA is:

> (DA) The RE depends on the DE, and if an RE depends on a DE, then the RE is superfluous. Therefore, the RE is superfluous.

One possible objection is simply to appeal our previous conclusion – since REs do add to our understanding, they are not superfluous, even if they are dependent. However, the dependency claim is worth looking at in more detail because doing so suggests an alternative account of the relation between REs and DEs.

To begin, the relevant notion of dependency appears to be 'order of construction', as illustrated in Chemero's methodology for constructing an RE for the HHC robot:

> It was [the] mathematical description of the control system in terms of feedback loops that allowed me to predict what behavior would be produced when particular patterns of activation were produced in the system. To find

[5] Due to space constraints I've removed further discussion of the PA. At one point Chemero appears to endorse the standard view that teleological explanations explain why a component is present in a system (Chemero, 2009, p. 71). If this is correct, then the issue is how to reconcile this endorsement with the subsequent rejection of the utility of teleological explanations. One possibility might be to grant that non-representational teleological explanations *do* add to our understanding of a system, but deny that representational explanations do. In other words, allow teleology, and perhaps even natural meaning, but draw the line at full-blown representations as defined by the received view. Such a position might then preserve the explanatory role of teleology while allowing for the rejection of representation as a useful explanatory tool. In reply, this seems like an ad hoc solution, for, if we allow functions to be explanatory, then why exclude one particular subtype of function from our explanatory toolkit?

> out what those patterns of activation represented, I determined what environmental situations those activations would adapt the agent to, in particular, by determining what environmental situations the ensuing behavior would be appropriate to. So the representational description is dependent upon the dynamical one. (Chemero, 2009, p. 77)

Let us acknowledge that some REs are dependent in the sense that they are constructed through consultation with a DE. This alone is not sufficient to justify the rejection of RE, however. Even if *dependent* REs were explanatorily superfluous, it may still be the case that *independent* REs are not. If the mark of superfluity is that an explanation is constructed through consultation with a DE, then those that are not so constructed may nonetheless be relevant.

This may explain why Chemero actually makes a much stronger claim about dependency. In particular, he writes that, for systems such as the HHC robot, "one *must* have the dynamical story first, before one can concoct a representational story" (Chemero, 2009, p. 73, italics added). The claim thus seems to be that, for a certain type of system, there are no REs absent a prior DE. I argue that this strong claim is probably unfounded in Chemero's test case – the HHC robot. The reason is that the DE is itself based on a prior *mechanistic explanation* (ME), and MEs are teleological.

Mechanistic explanations explain the behavior of a target system by specifying a set of components, their causal capacities, and their organization such that the behavior is manifested by that system (Craver, 2007; Bechtel & Richardson, 1993). MEs are teleological in the sense that they tell us what the components *do*, i.e., how they *function* in the context of the containing system. Given the received view, representational mechanisms are simply special cases where the function of a component is to carry information to a consumer. In other words, if we have a ME and information, we have the basic resources for representation.

The DE of the HHC robot is based on a prior ME. First, the *nodes* of the ANN controller are defined mechanistically in terms of their components ("channels") and how they interact to produce node-level behavior (HHC, 1995, p. 88). Second, the dynamical analysis is based on the control system for a single robot whose architecture is well defined. Figure 1 shows the mechanism of the robot, including the causal relationships that drive the movement of the robot given sensory inputs. But the moment we have the ME, we have the possibility that states of components in that mechanisms are representations – all that is required in addition is that those states play a certain functional role in producing system-level behavior, namely, the role of carrying information.

In fact, prior to undertaking the dynamical analysis, HCC simplify the network by making this very assumption. HHC look at the correlation between type of information being presented – visual or tactile – and the responses of hidden nodes, and remove those nodes "that play no part in visually guided behaviors " (1995, pp. 95-6) (figure 2). This process only makes sense if HHC are assuming that the function of some nodes is to carry visual information to later components.

To sum, prior to the analysis of the controller, there is (i) a decomposition of the controller into components, their activities, and their organization such that the behavior is manifested, and (ii) an assumption that the function of hidden nodes is to carry specific types of information to the motor nodes. Given the received view, we thus have the resources available for an RE of nodes in the network. Since the DE is based on a prior ME, and REs can be constructed off of MEs, it follows that an RE for the HHC robot need not depend on a prior DE.

However, there is another possible route to the strong notion of dependency favored by Chemero. An RE based on the ME alone will not be the *same* RE as a dynamically constructed RE. Since it is reasonable to suppose that some REs are better than others, one might argue that the dynamical-RE is *better* than the (mere) mechanistic-RE. The dependency claim can then be recast: Because a DE is the best possible way to predict the behavior of a certain type of system, the best RE *must* be constructed on the basis of that DE.

Two possible replies: First, further argument is required to guarantee that the dynamical-RE would be better, as this issue seems to be relative to our explanatory goals. In particular, it is not obvious that the dynamical-RE is always necessary to tell us *why* a component is present in the system. For example, determining that state *N* in a frog's brain has the function of carrying information about flies seems sufficient to explain the presence of *N*, and can be accomplished using traditional methods of intervention. Similarly, we know after HHC's process of simplification that the remaining hidden nodes are present because of their capacity to carry visual information.[6] A dynamical-RE may give us a more nuanced picture of what is represented, but that additional information may not always be necessary for the RE to fulfill its explanatory role.

That being said, the second reply is to concede that REs constructed on the basis of DEs *are* preferable to those constructed without. However, given the result of section 3, this does not show that the dependent RE is superfluous.

5. Conclusion: The Role of Dynamical Explanation in Representation

I've argued that (i) REs do add to our understanding of a system, and (ii) that REs are not always dependant on DEs, although (iii) I conceded that in some cases, REs based on DEs may be 'better' in some undefined sense. What might this sense be?

---

[6] One objection that comes to mind is this: The problem with systems that require explanation through dynamical systems theory is often that assessing the information processing role of an internal state is very difficult through traditional methods, e.g., by varying environmental stimuli and measuring internal state changes in the hope of finding 'detectors' tuned to external features easily describable in natural language terms (e.g., 'edge' detectors or the 'grandmother' cell). Instead, understanding information flow in such systems is greatly assisted by the tools of dynamical systems theory, and hence the only viable RE will be dependent on a DE. Given the argument of section 3, which preserves an explanatory role for representation, I am inclined to concede this point, with the following caveat: Since the Dependency Argument is *epistemological*, it must address the issue of whether there are cases where DE-independent REs for these systems can nonetheless satisfy the explanatory demands placed on them, as it is not obvious that there are not, as illustrated by HHC's use of traditional intervention methods to assess the functions of sub-networks of the robot controller.

It stands to reason that we want our REs to accurately reflect *how* and *what* a state represents, just as we want our DEs to accurately capture the causal structure of an agent-environment system. The suggestion, then, is that DEs may lead to a better understanding of representation, rather than displacing it.

For example, one important feature of DEs is that they are good at capturing how the states of components are tightly causally integrated through time. Let us distinguish between activities that take place *over* time and those that take place *in* time.[7] For example, 'writing a book' takes place over time as indicated by the fact one can truly assert that they are writing a book even if, at the time of making that utterance, they are not actually typing on their typewriter. In contrast, an activity such as walking takes place in time insofar as, if you pause while walking, you are no longer walking.

Traditional computational theory stipulates that computation is a process that happens *over* time, not *in* time – this is how a laptop computer can 'sleep' without disrupting the program that was running at the time of interruption.[8] Similarly, the received view is at least consistent with the claim that representing is an activity that takes place over time. For example, representational contents are typically presented in such a way that allows for 'freezing' the information processing in such a way that those processes could be restarted without fundamentally altering them.[9]

This, of course, is precisely where the contrast between traditional computational explanations and dynamical explanations is often drawn (Van Gelder, 1995). It is sometimes also assumed that this is where the representation / anti-representation divide occurs, yet as Chemero points out, this is incorrect: We can discard the commitment to traditional computation while keeping representational explanation, resulting in non-computational representational explanations (Chemero, 2009, p. 71). So, one possible way in which DEs may clarify our understanding of representation is by *helping us reconceptualize representation as an activity that is essentially temporal*. This, of course, is just a suggestion, but it points towards an alternative to Chemero's view about the relation between REs and DEs: Rather than working against each other, they work together.

[7] This distinction was reported to me through hearsay, and I was told that it originally appeared in an article published in a proceedings volume for an artificial intelligence conference, but I have been unable to secure a reference.

[8] Note that the pausing itself happens *in* time, not *over* time. So, for example, perhaps it is possible to pause the Watt governor by stopping time, but if we try to pause it *in* time, the momentum of the arms is lost (even if we hold the arms at the current angle), and starting it up again disrupts the information carried by the arms.

[9] Of course, if we pause an information-processing stream the information being processed will take longer to reach those areas that use that information, and hence the system may fail at a task it would have succeeded at had the stream not been paused (e.g., catching a fly). But the issue is here is whether pausing implies that the processing has been fundamentally altered, or whether it can start up again without incurring any errors.

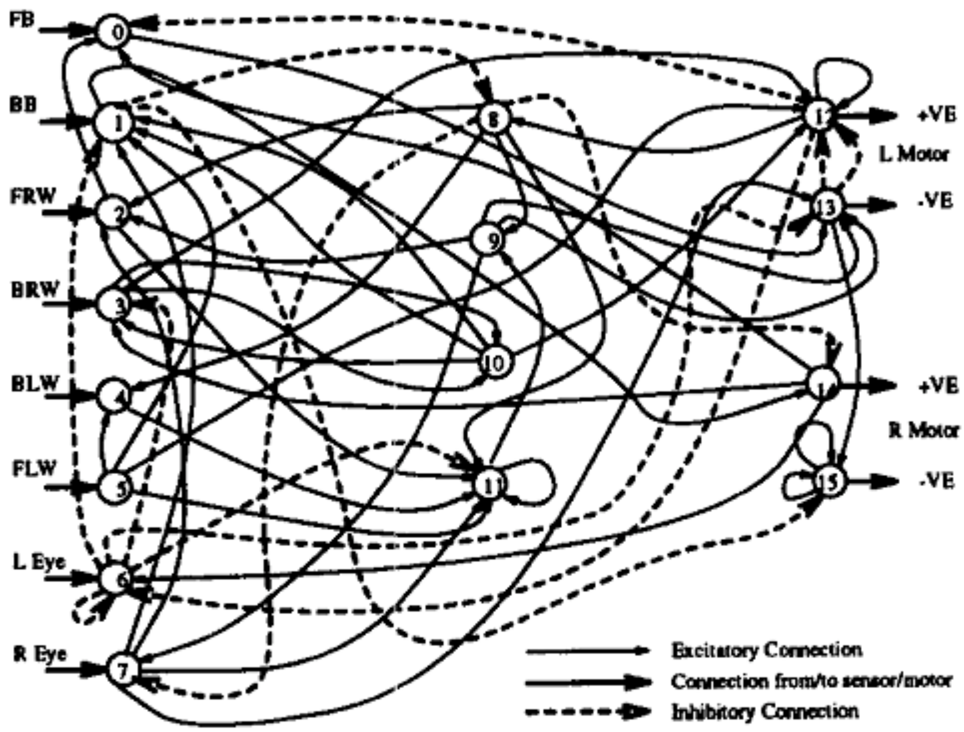

Fig. 11. Full C2 control network. The left-hand column are units originally designated as input units: FB=Front Bumper; BB=Back Bumper; FRW=Front Right Whisker; BRW=Back Right Whisker; BLW=Back Left Whisker; FLW=Front Left Whisker. Right-hand column shows output units, which are paired and differenced to give two motor signals in the range $[-1, 1]$ from four "neuron" outputs in the range $[0, 1]$. Centre column shows "hidden units".

Figure 1 – The full control network. (From HHC, 1995)

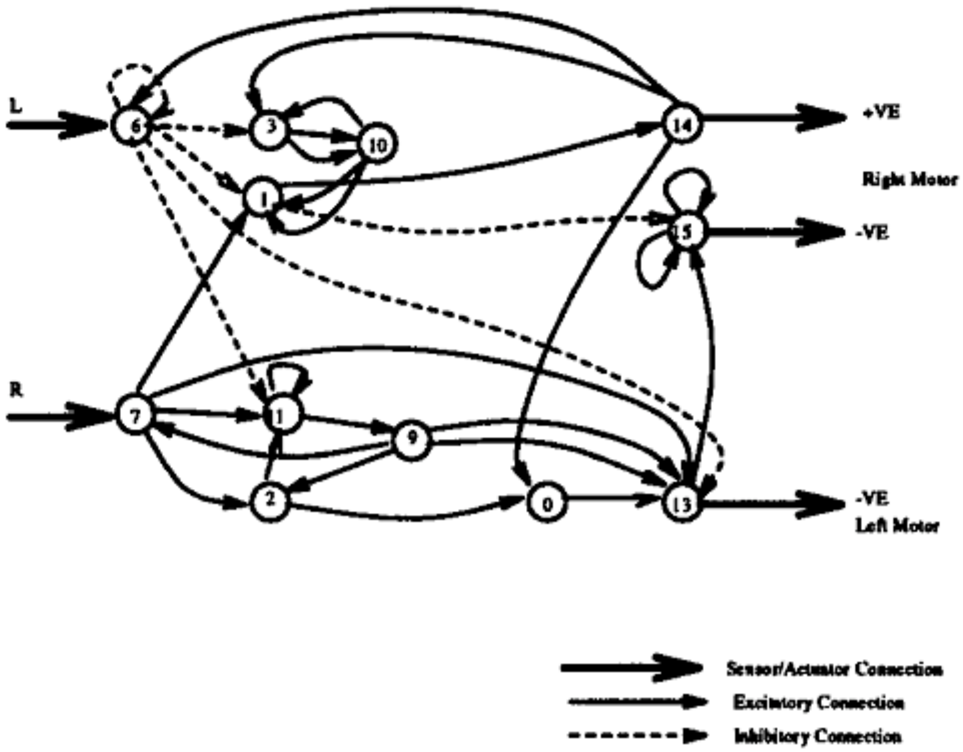

Fig. 12. C2 visual guidance pathways. Note that, for the sake of clarity, the positions of the left and right motor outputs have been interchanged.

Figure 2 – the simplified control network (From HHC, 1995).

Bibliography

Bechtel, W. (1998). 'Representations and cognitive explanations: assessing the dynamicist's challenge in cognitive science'. *Cognitive Science*, 22(3), pp. 295-318.

Bechtel, W. & Richardson, R. C. (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Princeton, NJ: Princeton University Press.

Canfield, J. (1964). 'Teleological explanation in biology.' *The British Journal for the Philosophy of Science*, 14(56), pp. 285-295.

Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge, MA: MIT Press.

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

Dretske, F. (1988). *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press

Eliasmith, C. (2003). 'Moving *beyond metaphors*: Understanding the mind for what it is.', *Journal of Philosophy*, 100(10), pp. 493-520.

Husbands, P., Harvey, I., & Cliff, D. (1995). 'Circle in the round: state space attractors for evolved sighted robots.' *Robotics and Autonomous Systems*, 15(1-2), pp. 83-106.

Millikan, R.G. (1989). 'Biosemantics'. *Journal of Philosophy*, 86, pp. 281-297.

Millikan, R.G. (1996). 'Pushmi-pullyou representations'. In *Mind and Morals*, L. May, M. Friedman, & A. Clark (Eds.), Cambridge, MA: MIT Press, pp. 145-162.

Van Gelder, T. (1995). 'What might cognition be, if not computation?'. *Journal of Philosophy*, 91(7), pp. 345-381.