Dynamical Systems, Emergence and Explanation

Joel Walmsley

University of Toronto

June 2003

1. A natural affinity?

The application of concepts from dynamical systems theory to issues in cognitive science is a growing and, I think, laudable project. Such usage has been extended from the neurophysiological level (e.g. Skarda and Freeman, 1987, Globus, 1992) through the intermediate "subsymbolic" level of Smolensky's connectionism (see also Horgan and Tienson, 1996), right up to the "personal" level of belief/desire psychology (e.g. Townsend and Busemeyer, 1995). Whilst not entirely uncontroversial, such applications have often yielded interesting insights, both empirical and theoretical.

One of the most intriguing issues concerns the concept of emergence, discussion of which often appears alongside the idea of dynamical approaches to cognitive science.[1] This is not entirely surprising. After all, the language of dynamical systems theory seems to be ideal for describing and explaining emergent properties. Many systems in artificial life, for example, display "emergent" behaviours which result from the collective and interactive action of a system's component parts, rather than from the action of any single component. This seems to be captured by the emphasis on "total state" explanations in dynamical systems theory to the point where dynamical language has even been vaunted as one of the "calculi of emergence" (Crutchfield, 1994).

There consequently seems to be a natural affinity between dynamics and emergence. But the intuition of this affinity is so strong that the connection between the two is seldom expounded in sufficient detail. Rather, we often encounter impressionistic or incomplete accounts of the relation between them. Van Gelder, perhaps the most well known dynamical cognitive scientist limits his treatment of emergence to the following:

---

[1] The locations of such co-occurrence are also interesting -- it certainly seems to be the case that more emergentists allude to dynamical systems theory than dynamicists allude to emergence. Explaining *why* that is the case is a matter for a different paper…

> "From a broadly dynamical perspective, cognition is seen as the emergent outcome of the ongoing interaction of sets of coupled quantitative variables rather than as sequential discrete transformations from one data structure to another" (p. 12)

Similarly, in an illuminating paper in which he discusses emergence and dynamical systems as they relate to connectionism and artificial life, Elman (1998) seems to dodge the very issue of the precise relation *between* emergence and dynamics. Instead, his comments on this topic are limited to:

> "...[A] theme common to much of the work in Alife is the emergence of structure and behaviours which are not designed, but rather the outgrowth of complex interactions.
>
> The dynamical systems approach is also concerned with interaction and emergentism; more generally it can be viewed as a mathematical framework for understanding the sort of emergentism and the high-order interactions which are found in both connectionist and artificial life models" (p.504)

That the dynamical systems approach is concerned with emergentism is perhaps uncontroversial -- here I am discussing the relation of these ideas, and alleging their frequent co-occurrence: a dynamicist (of sorts) concerning myself with emergence. What I want to focus on is a problem one might encounter in trying to say *how* these concepts are related -- it turns out, I think, that there are some barriers to constructing a dynamical framework for understanding emergence. If we don't question the apparent natural affinity then we are at risk of running into such barriers.

I shall have more to say about the concept of emergence in due course, but it is worth noting the conception of emergence that is prevalent in the kinds of cases that generate the appearance of a natural affinity. In the cases where emergence is invoked the "emergent" behaviours are often unanticipated (i.e. unplanned or unprogrammed). The conception of emergence, then, is in terms of *novelty* or *unpredictability.* It seems fair to say that this is the received view of emergence. In general, one is ill-advised to consult a dictionary for illumination on a philosophical matter, but nonetheless, the online version of the *Oxford English Dictionary* reflects this received view:

emergent, *a.* and *n.*

A.(*adj.*) 2.c. *Science.* That which emerges unpredictably as the result of an evolutionary process, *spec.* in ***emergent evolution***.

B.(*n.*) 3. *Science.* An effect produced by a combination of several causes but not being capable of being regarded as the sum of their individual effects. Opposed to *Resultant.*

For this reason, any property that could be *predicted* in advance of its occurrence by inspection of the parts of some system would not count as an emergent property of that system. The mass of a car, for example, is not an emergent product of the assembly line, since it is antecedently predictable from knowledge of the mass of all the parts. The property of consciousness, cognition, intentionality, or whatever happens to be a favourite "mark of the mental", so the story goes, is emergent, because one could not predict its precise form even if one were in complete possession of complete knowledge of all the properties of the brain from which it emerged.

Emergentists, then, must be opposed to the deductive-nomological explanation of emergent properties, since such a model allows for no distinction between prediction and explanation. According to the covering-law model of explanation, a phenomenon is explained when a statement describing it can be derived from statements specifying a law or set of laws and relevant initial conditions. Thus, an explanandum is, according to the model, a *deductive consequence* of an explanans. The only difference between a prediction and an explanation is whether or not the state of affairs described in the explanandum is known to have obtained. It is therefore solely a pragmatic difference -- prediction and explanation have an identical logical structure, but differ in terms of what one knows and what one wants to know. On the one hand, you can predict some event before it happens if and only if you could explain it after it happens. On the other hand side you cannot explain an event if and only if you could not have predicted it.

Accordingly, if one adopts a covering-law model of explanation, the phrase "explaining emergence" takes on an air of oxymoron. If one could explain some emergent property or event, E, one would also be able to predict E, hence E would not be emergent *qua* unpredictable. If E can be *deduced* from the conjunction of a statement of a law or set of laws, $L_1 ... L_n$, and a statement of initial conditions, C, then E cannot be emergent *qua* unpredictable, since it follows *a priori* (i.e. as a deductive consequence of $L_1 ... L_n$ and C). I suspect that it is this line of thought which led Samuel Alexander (1920), in the heyday of British emergentism, to write:

> "The existence of emergent qualities thus described is something to be noted, under the compulsion of *brute empirical fact* [...] *It admits no explanation*" (p.46, my emphasis)

2. Dynamical Explanation

Consequently, a nagging question which must be answered concerns what kind of explanation dynamical explanation actually *is*. The official story (told, for example, by Van Gelder) is that dynamical explanation is a new form of explanation, to be contrasted with more familiar styles of explanation in cognitive science. In a dynamical explanation, such as those found in some Smolensky-style connectionist work, the state or behaviour of a system is to be explained with reference to points and trajectories in an abstract state space, where that space has one dimension for each variable of the system.

More familiar styles of explanation in cognitive science are sometimes called "morphological" and "systematic" explanations. Here, the ability of a system is explained by the abilities of its parts or the interactions of its parts. Given that, in some cases, variables describing a dynamical system do not refer to structural features of the system being explained (they may, for example, be collective parameters), we cannot consider dynamical explanation a form of morphological or systematic explanation. As Andy Clark (2001) puts it, dynamical explanation "differs from more traditional cognitive scientific explanations in that it greatly abstracts away from the behaviour of individual systemic components"(p.124)

This abstraction away from physical details has led some to object that, in fact, dynamical explanations amount to little more than descriptions of the data -- that they are, at best, an exercise in curve fitting, but do not explain *why* the data take the form they do. There are two, I think satisfactory, answers to this objection. The first, outlined by Van Gelder (1998) is one from analogy. Dynamical explanations of cognition are, he claims, of a kind with, for example dynamical explanations of celestial motion. Since the latter count as paradigm examples of scientific explanation, there is no reason to regard the former as somehow defective. Of course, a *poor* dynamical explanation could amount to no more than a curve-fitting description, but the problem with such an account would be that it was poor, not that it was dynamical.

The second answer to this objection points us towards an understanding of what dynamical explanations actually are. Dynamical explanations are more than descriptions of the data at least in the sense that they support counterfactual reasoning. In other words, a good dynamical explanation will enable us to say how the system would have behaved in various nonactual circumstances -- for example if it suffered specific perturbations, or if its control parameters were altered.

Bechtel (1998) takes this ability of dynamical explanations as evidence of their being deductive-nomological explanations:

> "DST accounts, such as [Townsend and Busemeyer's (1995) Decision Field Theory] are clearly designed to support counterfactuals. They are designed to tell what would happen under different motivational values, for example. This suggests that it may be appropriate to construe these DST explanations as being in the covering law tradition" (p.311)

It is worth taking some time to strengthen this point since the official view espoused by the likes of Van Gelder is that dynamical explanations are *not* covering-law explanations. He rightly notes that in some case, dynamical explanations can be provided without the citation of a general law of physics. According to the strictest deductive-nomological model of explanation, $L_1 \ldots L_n$ cannot be *any* old laws -- they must be basic laws of physics. In a dynamical explanation, the law which gets cited is an equation for the path of the trajectory through state space. One generally supposes that such dynamical equations have their ground in laws of physics -- indeed in some cases, though as yet none in cognitive science, it may be shown how these dynamical equations can be derived from the laws of physics. But of course the absence of such a derivation does not rob a dynamical explanation of any of its explanatory force, since the derivation of a dynamical law from a law of physics is not *itself* part of the dynamical explanation.

I am prepared to concede the point that dynamical explanations are not strictly deductive-nomological models in the sense which has become important in philosophy of science. But dynamical explanations are still a distinct species of *covering law* explanation where the general law under which the specific case falls is not a law of physics, but a (higher level) dynamical law. The problem is that even this weakened type of covering law explanation is incompatible with any conception of emergence in terms of novelty, unpredictability or non-deducibility. Whether it's a basic physical law, or a higher level dynamical law, the final state is still *deducible* from the statement of that law together with a statement of initial conditions and so cannot be considered novel or unpredictable.[2]

Some quotations should help to illustrate this point. Van Gelder (1991) himself outlines the *goal* of the dynamical approach to cognition, writing:

[2] I leave to one side the issue of deducibility in systems which obey stochastic laws. I suspect, however, that such laws could still facilitate a probabilistic prediction of final states in accordance with Van Gelder's "goal" below.

> "In studying and explaining the behaviour of dynamical systems one aims at formulating equations which describe the evolution of the system, and can be consequently used to explain why the system is in the state it is in, or to predict what states it will come to be in" (p.500)

He later goes on specifically to emphasise the *deductive* nature of dynamical explanations:

> "If we know the current state of the system -- i.e. the point in state space it currently occupies -- we can use the equations governing the behavior of the system to determine what point it will occupy next" (p.500)

This view on deducibility is reiterated in a deterministic form in Van Gelder and Port (1995):

> "Dynamical modeling ... involves finding ... a mathematical rule, such that the phenomena of interest unfold in exactly the way described by the rule" (p.14)

The same authors later draw our attention to the fact that knowledge of both the covering laws and the initial states is necessary for explanation-cum-prediction:

> "Dynamical models based on differential equations are the pre-eminent mathematical framework science uses to describe how things happen in time. Such models specify how change in state variables at any instant depends on the current values of those variables themselves and other parameters. Solutions to the governing equations tell you the state that the system will be in at any point in time, as long as the starting state and the amount of elapsed time are known" (p.19)

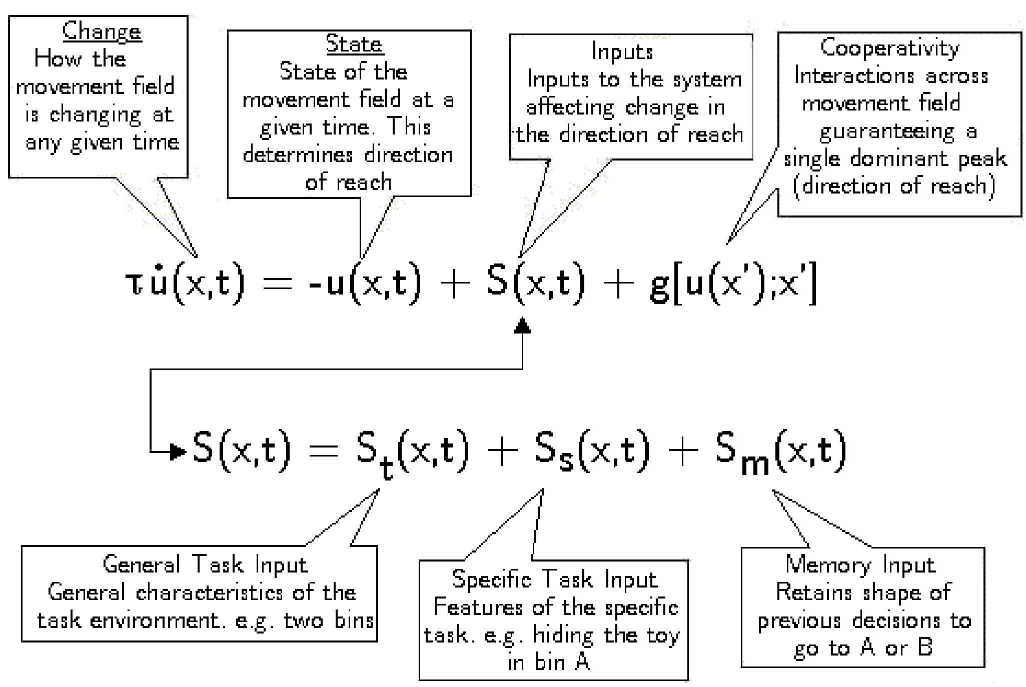
I think that such considerations are sufficient to establish that, even if dynamical explanations are not strictly deductive-nomological, their character as special cases of *covering-law* explanations is still strong enough to rule out emergence.

The goal of dynamical cognitive science, as Van Gelder succinctly describes it, appears to have been met by the successful models which dynamicists in cognitive science have constructed. The point that dynamical explanations are covering law explanations is thus borne out by example -- it appears that such models are the quest to find the laws of thought, in the form of differential equations. For illustrative purposes, let me focus on the model put forward by Thelen, Schöner, Scheier and Smith (2001) to explain the classic "A-not-B" reaching error in infants. Piaget first observed that if you take an infant, between 7 and 12 months old and hide a toy he likes under one of two bins in front of him, he will reach for the correct bin in order to get the toy. However, if you hide the toy under bin A a few times, and then switch to hiding the toy under bin B, the infant makes the A-not-B error and reaches towards bin A.

Why does this happen? Piaget postulated that it was something to do with the infant's understanding of object permanence. The main effect is quite reliable, but it is enormously sensitive to slight changes in the experimental conditions, such as the delay between viewing and reaching, the way the scene is viewed, the number of trials and so on. Some have contended then that the error arose because of the infant's underdeveloped concept of space (ego- versus allocentrically represented). Others have suggested that the infant's *memory* is to blame.

Each of the hypotheses manages to explain a limited subset of the data, but fails to deliver an adequate account of the overall phenomenon. Thelen et al. contend that this is because all of the hypotheses focus on the infant's *internal* machinery -- the way he thinks about the world. They suggest, instead, that we focus on the infant's *reaching activity*, and have accordingly developed a dynamical model of how the infant comes to reach in a particular direction. They conceive of a number of factors including memory of previous reaches, motivational level, characteristics of the task domain and so on, as parameters which can be quantified and related in a "grand equation" as in figure 1.

Figure 1: Thelen et al.'s (2001) "Grand Equation", (Summarised in Van Gelder, 1999)



The grand equation specifies how the infant's inclination to reach in a certain direction changes over time as a result of changes in these parameters. What's most impressive is that the resulting dynamical system reproduces the classic A-not-B error, together with all of the contextual subtleties which caused problems for previous hypotheses, but it does so without making reference to inner representation-level rules or algorithms.

By emphasising the "grand equation", Thelen and Smith offer what must be regarded as a covering-law explanation. To explain the infant's behaviour, all you need to do is cite the grand equation together with the initial values of the seven parameters of the grand equation and *voila*: the infant's behaviour, also expressed in terms of different values of the seven parameters, follows as a deductive consequence. But this is bad news for any claims to the A-not-B behaviour being emergent *qua* novel. The infant's reaching behaviour is entirely predictable in advance of its occurrence.

3. Reconceiving Emergence

We have, then, a tension. Not just a contrived philosopher's tension, but a genuine tension which arises in the literature when, as is frequently found, dynamical explanation and emergence are referred to in the same breath, and the former is alleged to illuminate the latter. It is worth mentioning that dynamical cognitive scientists such as Van Gelder and Bechtel do not run into this problem, since they do not concern themselves with the question of emergence. The problem arises for those who (like myself, I'm prepared to admit) want to make room for both emergence and dynamical systems in cognitive science. The question then becomes "Is there a way to sustain the intuition of a natural affinity between dynamical systems theory and emergence, or must we, in the light of the tension between them, let one go?"

Before we answer this question we should consult the dialectical road-map and note that there are clearly two potential paths of escape from the apparent tension. One is to deny that dynamical explanations are covering-law explanations, the other is to deny that emergence is, in fact, precluded by covering law explanations. The considerations of the previous section have, I think, blocked off the first path -- dynamical explanations, even if they're not strictly deductive-nomological, are nonetheless a form of covering law explanation strong enough to rule out emergence *qua* novelty or unpredictability. This is borne out by statements of the dynamicist goal, as well as paradigm examples of (successful) dynamical models.

This leaves one remaining path of escape from the tension. We could attempt to conceive of emergence in a way that does not clash with the covering-law character of dynamical explanation. Though never set up explicitly as an escape from the tension I have outlined, the strategy of reconceiving emergence is not without precedent in the literature.

Clark (1997) discusses the notion of emergence as it applies to embodied cognition, both "natural" and "artificial". He argues, against the received view as I have called it, that the characterisation of emergence in terms of unexpected or unpredictable features or behaviours is inadequate. This is because unexpectedness or unpredictability is overly observer-relative -- after all, what might be predictable for one person (say, a robotics engineer), might be unexpected for another (say, anyone who is not a robotics engineer). What, he claims, is needed, is a "criterion less hostage to the vagueries [sic] of individual expectations" (p.109).

Clark's criticism of emergence *qua* unpredictability in fact misses the mark, for it depends on an equivocation over the sense of "predictable". What Clark is criticising is a definition of emergence in terms of *psychological* predictability. It is hard to disagree with his point when predictability is observer-relative, after all, it seems likely that even the most obvious cases of non-emergence could be unpredictable for *someone*.

But what is at stake, I take it, is a notion of emergence in terms of *logical* predictability, or "deducibility". We can emphasise the distinction by considering any problem from a logic textbook (this one is from Hodges' *Logic*). A person, say a new student, who is not well versed in the predicate calculus would not be able to predict $\forall x[x = b \leftrightarrow [Cx \wedge Bx]]$ from the conjunction of $\forall x[x = b \leftrightarrow [Mx \wedge Cx]]$ and $\forall x[Cx \rightarrow [Bx \leftrightarrow Mx]]$. In this case, $\forall x[x = b \leftrightarrow [Cx \wedge Bx]]$ is not *psychologically* predictable for the student. It may be, of course, psychologically predictable for some other person, say the professor. But none of these "vagaries of individual expectancies" impinge on the fact that $\forall x[x = b \leftrightarrow [Cx \wedge Bx]]$ is *logically* predictable (i.e. deducible) from the conjunction of $\forall x[x = b \leftrightarrow [Mx \wedge Cx]]$ and $\forall x[Cx \rightarrow [Bx \leftrightarrow Mx]]$. The student would be ill advised to claim that she did not know the answer because the answer was emergent. But were the conclusion *not* logically predictable from a conjunction of premises, it would not be the sort of example that would appear in a logic textbook.

Thus, the absence of *psychological* predictability is not, as Clark rightly contends, the hallmark of emergence. But psychological predictability is not the only form of predictability. It turns out that it is the of absence of *logical* predictability that has often been associated with emergence. Indeed, this view is the one that is present in some of the canonical emergentist texts. Broad, for example, in his (1925) *Mind and Its Place in Nature,* sees emergence as dependence and determination without *deducibility* -- an emergent property of some whole, w, is one that depends on and is determined by the properties of the parts of w, without being deducible from the properties of the parts of w. Further, this view is still present in the contemporary literature. Van Cleve (1990) for example, defines emergence in these terms:

"If P is a property of w, then P is emergent iff P supervenes with nomological necessity, but *not* with logical necessity, on the properties of the parts of w" (p.222)

A definition of emergence in terms of the absence of *logical* predictability does it fact fit Clark's desideratum -- it is a criterion that is entirely free of the vagaries of individual expectations. His criticisms of emergence *qua* unpredictability do not, therefore, hold water.

Regardless, Clark goes on to give a different definition of emergence whereby "a phenomenon is emergent if it is best understood by attention to the changing values of a collective variable" (p.112) It follows, he notes, that "Emergence, thus defined, is linked to the notion of what variables figure in a good explanation of the behaviour of the system". In other words, this definition of emergence is *wholly* compatible with covering-law explanations of the sort that Van Gelder seeks and Thelen et al. provide, especially given that, as we noted above, the variables in covering-law dynamical explanations often refer to collective parameters, and thus take place at "one level of remove from the actual mechanisms which produce the behaviour quantified and explained in the dynamical account" (Van Gelder, 1991, p.501).

So we might be tempted to conclude that the problem has been overcome -- by following in Clark's footsteps, it turns out that there is a way to conceive of emergence which renders it compatible with the covering-law nature of dynamical explanations. A matter of some concern here, though, is the relation between the different extant conceptions of emergence. What matters for my present concerns is not which definition of emergence is the correct one, but rather that, in formulating these definitions, we have a situation where two sides are simply talking past each other. Those interested in the (for want of a better word) scientific import of emergence (à la Clark) and those interested in the metaphysical aspect of emergence are simply talking about *different things*. The former group, who consider emergence by paying attention to the changing values of a collective variable, can be entirely satisfied with covering-law type dynamical explanations. The latter group, who consider emergence in terms of novelty or unpredictability of system behaviours or states, must avoid such explanations and seek a different account if they want to avoid the charge of mystery implicit in the quotation from Alexander above.

This is a pretty devastating metatheoretical result for anyone who had hoped that scientific research on dynamical systems would help to illuminate metaphysical issues to do with the "emergence" of life and cognition. It's particularly problematic given that the two groups I have distinguished previously thought that they were talking about the same thing. It turns out that there are at least two extant concepts of emergence, these two concepts are incompatible, both philosophical and scientific research have failed to distinguish them, and interdisciplinary attempts to illuminate them are doomed to failure. The problem of emergence has not gone away.

4. Concluding Remarks.

By way of concluding, I'd like to formulate a call for help, as well as suggesting one direction from which I think that such help could come. Unlike some philosophers I don't enjoy unsolved problems, or revel in deflationary arguments. With little in the way of justification, I should say that I find emergence and dynamical systems to be quite intuitively appealing as ways of helping to understanding cognition. Moreover, I like to think that philosophy and science can sometimes help each other out in quite a fruitful way. At the moment, it looks as if it will turn out that that intuition, as is the case with many intuitions, is in fact unwarranted.

Space and time prevent me from exploring it more fully, but there may be a way in which the emergentist-dynamicist can at least partially dig themselves out of the hole I've created. Recall what I was saying earlier about *derivation*. It's quite clear that if you can derive a statement of final conditions from the conjunction of statements of a law and a statement of initial conditions, then there can be no emergence. But I also mentioned that a widespread assumption is that dynamical equations will ultimately be derivable from the laws of physics, even if no such derivation is, as yet forthcoming. There could conceivably be cases, however, where the dynamical covering law which features in a dynamical explanation is itself emergent in the sense that it cannot be derived from a law of physics.

In this case, it would turn out that any behaviour of the system could be predicted on the basis of some higher-level law, but that this higher-level law could not be predicted from statements about (say) neurophysiology or physics. In this case, the emergence would have, as it were, happened before the dynamical explanation got started -- we would have a kind of emergence which was "nomological" in the sense that it was the *law itself* which emerged.

So, from this impressionistic sketch, the problem may not be insurmountable. It may turn out that there is a way to conceive of emergence such that it is *compatible* with dynamical covering-law explanations. The down side to this account is that we can no longer expect dynamical systems theory to be, in Elman's words, a mathematical framework for *understanding* emergence. Thus, the best-case scenario is that at least dynamical explanations need not *rule out* there being emergent laws, but (unfortunately) neither will they *account* for that emergence.

5. References

Alexander, S. (1920) *Space, Time and Deity* (London: Macmillan and Co.)

Bechtel, W. (1998) "Representations and Cognitive Explanations: Assessing the Dynamicist's Challenge in Cognitive Science" *Cognitive Science* 22 295-318

Broad, C.D. (1925) *Mind and Its Place in Nature* (London: Routledge and Kegan Paul)

Clark, A. (1997) *Being There: Putting Brain, Body and World Together Again* (Cambridge, MA: MIT Press)

Clark, A. (2001) *Mindware: An Introduction to the Philosophy of Cognitive Science* (Oxford: Oxford University Press)

Crutchfield, J.P. (1994) "The Calculi of Emergence: Computation, Dynamics and Induction" *Physica D* 75 11-54

Elman, J.L. "Connectionism, Artificial Life and Dynamical Systems: New approaches to old questions" in W. Bechtel and G. Graham (Eds.) *A Companion to Cognitive Science* (Oxford: Blackwell)

Globus, G. (1992) "Toward a noncomputational cognitive neuroscience" *Journal of Cognitive Neuroscience* 4 299-310

Hodges, W. (1977) *Logic* (London: Penguin)

Horgan, T., and Tienson, J. (1996) *Connectionism and the Philosophy of Psychology* (Cambridge, MA: MIT Press)

Skarda, C.A., and Freeman, W. (1987) "How brains make chaos in order to make sense of the world" *Behavioral and Brain Sciences* 10 161-195

Thelen, E., Schöner, G., Scheier, C., and Smith, L.B. (2001) "The dynamics of embodiment: A field theory of infant perseverative reaching" *Behavioral and Brain Sciences* 24 1-86

Townsend, J.T. and Busemeyer, J. (1995) "Dynamic Representation of Decision-Making" in R. F. Port and T. Van Gelder (Eds) *Mind as Motion: Explorations in the Dynamics of Cognition*

Van Cleve, J. (1990) "Mind-dust or magic? Panpsychism versus emergence" *Philosophical Perspectives* 4 215-226

Van Gelder, T. (1991) "Connectionism and Dynamical Explanation" in *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (Erlbaum)

Van Gelder, T. (1998) "The Dynamical hypothesis in cognitive science" *Behavioral and Brain Sciences* 21 615-665

Van Gelder, T. (1999) "Revisiting the Dynamical Hypothesis" Preprint No. 2/99, University of Melbourne, Department of Philosophy. Available at http://www.arts.unimelb.edu.au/~tgelder/papers/Brazil.pdf

Van Gelder, T. and Port, R.F. (1995) "It's about time: An Overview of the Dynamical Approach to Cognition" in R. F. Port and T. Van Gelder (Eds) *Mind as Motion: Explorations in the Dynamics of Cognition* (Cambridge, MA: MIT Press)