Revisiting the Dynamical Hypothesis

Tim van Gelder
Department of Philosophy
University of Melbourne
Parkville VIC 3052
Australia
tgelder@ariel.unimelb.edu.au

> "There is a familiar trio of reactions by scientists to a purportedly radical hypothesis: (a) "You must be our of your mind!", (b) "What else is new? Everybody knows *that*!", and, later—if the hypothesis is still standing—(c) "Hmm. You *might* be on to something!" ((Dennett, 1995) p. 283)

1. Introduction

Here are some claims about cognitive science which, it seems to me, no sane person could deny:

1. Increasingly, cognitive scientists are using dynamics to help them understand a wide range of aspects of cognition.[1]
2. Dynamical systems theory and orthodox computer science are rather different disciplines; typical dynamical systems and typical digital computers are rather different kinds of things.
3. Dynamically-oriented cognitive scientists see themselves as understanding cognition very differently from their mainstream computational cousins.

These claims are my starting point. When I say that they are undeniable, I don't mean to pretend that they are unproblematic. Indeed, for a philosopher of cognitive science, claims like these really just set up a challenge. What is going on here? What exactly is dynamical cognitive science, and how, more precisely, does dynamical cognitive science relate to orthodox computational cognitive science? And once we have reasonable answers to questions like these, even more interesting questions arise. What are the prospects for dynamical cognitive science? And what does it tell us about the nature of mind? Is the mind really a dynamical system rather than a digital computer?

In a series of papers, I have taken a particular approach to addressing this broad cluster of issues. That approach begins with the observation that the essence of the mainstream computational approach to cognition is often taken to be encapsulated in Newell & Simon's "Physical Symbol System Hypothesis," the claim that "a physical system has the necessary and sufficient means for general intelligent action" (Newell & Simon, 1976). In the quarter-century since Newell & Simon put this hypothesis on the

---

[1] For a representative sampling, see (Port & van Gelder, 1995).

table, a lot of work has been done elaborating and articulating the core idea, and it is now more aptly expressed in the slogan that "cognitive agents are digital computers." But if this slogan captures the essence of the mainstream computational approach to cognition, then the obvious parallel for dynamical cognitive science is the slogan "cognitive agents are dynamical systems." The philosophical challenge is then to say what the dynamical slogan *means*, in a way that does justice not only to the key concepts but also to cognitive science as it is actually practiced "in the field."

My basic stand on the meaning of the dynamical hypothesis (DH), and its place in cognitive science, was laid out in a paper which appeared recently in *Behavioral and Brain Sciences* (van Gelder, 1998b). However the task of articulating broad, deep ideas about the nature of a whole discipline—especially a discipline like cognitive science, which seems to be in a constant state of flux—is one that can never be definitively completed. Just as cognitive science is constantly evolving, so our philosophical understanding of the nature of cognitive science must also evolve. What I would like to do here is consider some of the most interesting objections to the dynamical hypothesis, as formulated in the BBS paper, and to consider how that formulation should be defended or adapted (Section 4). Before doing that, however, I will try to convey the flavor of dynamical cognitive science with a couple of illustrations (Section 2), and then present a précis of the basic dynamical hypothesis (Section 3).

Roughly speaking, you can tell dynamicists in cognitive science by the fact that their models are specified by differential or difference equations rather than by algorithms. However, describing the difference this way does little to convey the depth and flavor of the contrast between dynamical cognitive science and its orthodox computational counterpart. In my experience the easiest way to do this is to begin with an example, one that is drawn not from cognitive science but from the history of the steam engine.

2. Computational versus Dynamical Governors

Imagine it is sometime in the latter part of the 18th century, and you need a reliable source of power to drive your cotton mills. The obvious choice is the newly-developed rotary steam engine, in which the back-and-forth motion of a steam piston is converted to the rotary motion of a flywheel, which can then power your machines. The problem, however, is that for quality output your machines need to be driven at a constant speed, but the speed of the rotary steam engine fluctuates depending on a range of factors such as the temperature in the furnace and the workload. So here is your engineering problem: design a device which can regulate or "govern" the engine so it runs at constant speed despite the myriad factors causing variation.

The best way to control the speed of a steam engine is to adjust the throttle valve, which controls the amount of steam entering the piston. So the challenge becomes that of figuring out when, and by how much, to adjust the valve. From the vantage point of classical cognitive science, the proper approach seems obvious—attach a mechanism carrying out the following little algorithm:

1. Measure the speed of the flywheel;
2. Compare the actual speed against the desired speed;
3. If there is no discrepancy, return to step 1; otherwise
    a. Measure the current steam pressure.
    b. Calculate the desired alteration in steam pressure.
    c. Calculate the necessary throttle valve adjustment.
4. Make the throttle valve adjustment.

Return to step 1.

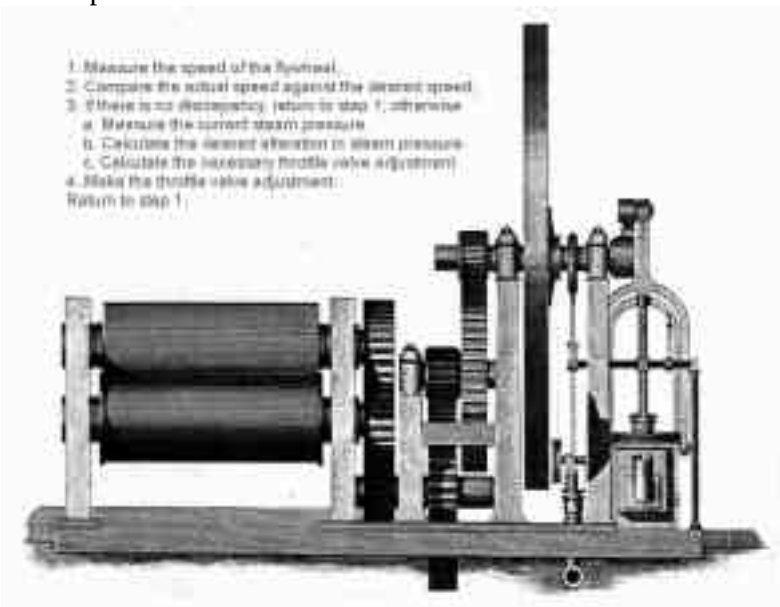

Figure 1: Steam engine driving a sugar mill. Quality output required constant speed from the engine.

Note that this mechanism, which we can call a computational governor, has to first take input in the form of measurements which result in symbolic representations of various aspects of the engine; then compute various quantities by manipulating symbols according to quite complex rules; and then convert the resulting specifications into actual throttle valve adjustments. The system is thus *cyclic*, *(digital) computational*, and *representational* in its design.

An additional subtlety is worth noting: the only *timing* constraint on the operation of the computational governor is at the output, where the throttle valve adjustments are made. These must be made sufficiently often to control the speed within acceptable limits. Within that frame, all other operations can happen at any time and at any speed. In this sense, timing within the computational governor is arbitrary; in other words, the device is in an interesting way *atemporal*.

Now, no doubt today it would be possible to build a governor working in this familiar computational fashion. However this was not the way the problem could have been solved in the eighteenth century. Most obviously, digital computers capable of handling the relevant calculations wouldn't be invented for another 150 years. The actual solution, developed initially by the Scottish engineer James Watt, was marvellously simple. It consisted of a vertical spindle geared into the main flywheel so that it rotated at a speed directly dependent upon that of the flywheel itself. Attached to the spindle by hinges were two arms, and on the end of each arm was a metal ball. As the spindle turned, "centrifugal" force drove the balls outwards and hence upwards. By a clever arrangement, this arm motion was linked directly to the throttle valve. The result was that as the speed of the main wheel increased, the arms raised, closing the valve and restricting the flow of steam; as the speed decreased, the arms fell, opening the valve and allowing more steam to flow. The engine adopted a constant speed, maintained with extraordinary swiftness and smoothness in the presence of large fluctuations in pressure and load.

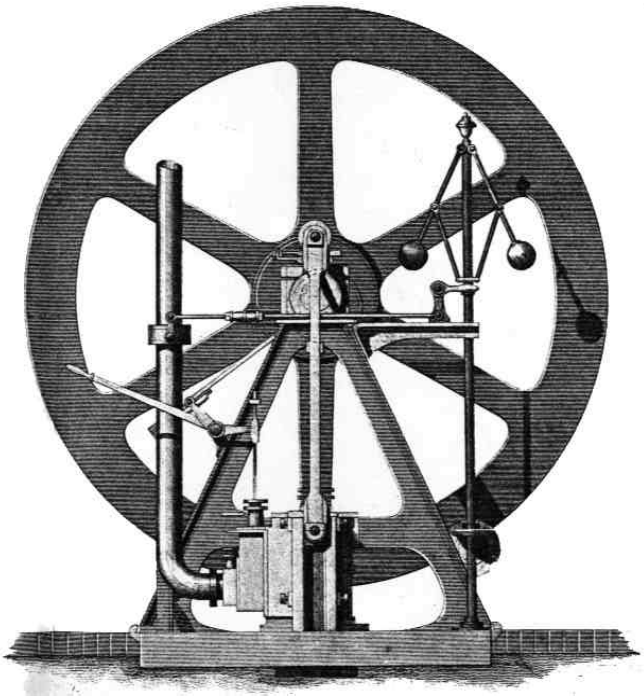

Figure 2: The Watt governor for controlling the speed of an engine;
also known as the centrifugal or dynamical governor.

It is worth emphasizing how remarkably well the Watt governor actually performed its task. This device was not just an engineering hack employed because computer technology was unavailable. In 1858 *Scientific American* claimed that an American variant of the basic Watt governor, "if not absolutely perfect in its action, is so nearly so, as to leave in our opinion nothing further to be desired."

The Watt governor is often known as the *centrifugal* governor, but it would be more accurate, and in the current context more convenient, to describe it as the *dynamical* governor, for the Watt governor is a classic example of a dynamical system as studied in dynamics textbooks. The key variable is the angle of the arms, θ (theta), whose behavior is described by the differential equation

$$\frac{d^2\theta}{dt^2} = (n\,\omega)^2 \cos\theta \sin\theta - \frac{g}{l} \sin\theta - r\frac{d\theta}{dt}$$

where n is a gearing constant, ω is the speed of engine, g is a constant for gravity, l is the length of the arms, and r is a constant of friction at hinges. This nonlinear, second order differential equation tells us the instantaneous *acceleration* in arm angle, as a function of what the current arm angle happens to be (designated by the *state variable* θ), how fast arm angle is currently changing (the derivative of θ with respect to time, $d\theta/dt$) and the current engine speed (ω). In other words, the equation tells us how change in arm angle is changing, depending on the current arm angle, the way it is changing already, and the engine speed. Note that in the system defined by this equation, change over time occurs only in arm angle θ (and its derivatives). The other quantities (ω, n, g, l, and r) are assumed to stay fixed, and are called *parameters*. The particular values at which the parameters are fixed determine the precise shape of the change in θ. For this reason, the parameter settings are said to fix the *dynamics* of the system.

Figure 3: James Watt, developer of the centrifugal governor.

In normal operation, of course, the dynamical governor is connected to the engine, which can also be described as a dynamical system. The two systems are said to be coupled, in this precise sense: the key variable in the engine system is its speed, ω, which is a parameter in the pendulum system, and the key variable in the pendulum system, θ, is a parameter in the engine system. When all is working as it should, the coupled engine/governor system has a stable fixed point attractor which is the desired constant speed.

The important lesson here is this: the dynamical governor is patently very different from the computational governor. Instead of cycles of inputs, symbolic representations, rule-governed, atemporal computations, and outputs, we have the continual mutual influencing of two quantities. This influencing is very subtle (though

mathematically describable): the state of one quantity is continually determining how the other is accelerating and vice versa. This relationship is very unlike the relationship between a digital symbol and its referent.

Now, let me be the first to point out that the dynamical governor is not *cognitive* in any very interesting sense. It is invoked to convey the general flavor of dynamical systems and how they can interact with their "environments," and for cognitive scientists, to spark the imagination and get the conceptual juices flowing. To really understand the distinctive character of dynamical cognitive science, we need to turn to the dynamical models themselves. As it happens, some of these models do actually bear a striking similarity to the dynamical governor / engine arrangement. One good example is a model recently developed by Esther Thelen and colleagues to account for that perplexing developmental phenomenon, the so-called "A not B" error.

3. Why Do Infants Reach to the Wrong Place?

The classic A not B error, as originally discovered and described by Piaget, goes something like this. Suppose Jean is an infant roughly 7-12 months old. At this age Jean knows what he likes—e.g., a toy—and when he sees it he will reach out for it. If you hide the toy in one of two bins in front of him, Jean will reach towards the right bin. But if you hide the toy in bin A a few times, and then hide it in the bin B, after a few seconds poor Jean makes the A not B error; he reaches towards bin A.

Why does this happen? Piaget's original explanation was cast in terms of the infant's emerging concept of an object. Jean is at "Stage IV" in this process, where infants seem to believe that an object has lasting existence only where it first disappeared. Call this the **cognitivist** explanation: the error is due to limitations in Jean's *concepts*.

The A not B error is fascinating because although the main effect is quite reliable in the standard setup, it is very sensitive to many kinds of changes in the experimental conditions. Many contemporary developmental psychologists reject Piaget's explanation, and its descendants, because it seems unable to account for Jean's reaching behavior under these alternative conditions. They have come up with at least two other major kinds of explanations.

According to the **spatial** hypothesis, for example, Jean's concept of an object is OK; he just has trouble moving his arm to the right place, and this is because he represents *space* wrongly. In this transitional stage Jean is still representing the world "egocentrically" rather than "allocentrically." Jean reaches in the direction the object usually is in relation to him rather than to where it now is in independent 3-D space. Thus if Jean is rotated to the other side of the table, he will now reach for bin B, which now occupies the "A" position relative to him. The **memory** hypothesis also maintains

that Jean's concept of an object is OK; however in this story he has trouble remembering where the object has been hidden. The memory is necessary in order to overcome the habit of reaching towards A. The memory hypothesis can account for why Jean actually does reach towards B if he is allowed to reach in the first few seconds after hiding. It is only later that he makes the error.

The cognitivist, spatial and memory hypotheses are ingenious attempts to explain a rich and perplexing body of experimental data. These explanations are very broadly similar to the computational approach to the governing problem, in that they focus attention on Jean's internal cognitive machinery, the way he *thinks* about the world. Also, while each one seems to capture *some* truth about the A not B error, none delivers an adequate account of the overall phenomenon. In each case there are some aspects of the experimental data the hypothesis cannot explain.

Thelen *et al* have taken a very different approach to the A not B error. Instead of focusing on the contents of Jean's *mind*, they focus on Jean's *reaching activity*. They develop a general, high level dynamical model of how we come to reach in a particular direction, and then explain the A not B error by applying the general model to the special circumstances of an infant at approximately 7-12 months.

Think of it this way. Suppose we are trying to choose a direction to reach in, with options ranging from far left to far right. Suppose also that our level of inclination to reach in any particular direction is constantly changing. In the Thelen et al model, this constantly changing set of inclinations is what they call the *movement planning field*, and is specified by a function u(x,t), which tells us our inclination to reach in direction x at time t. The heart of their model is a differential equation specifying how u is changing at any given time, depending on a number of further factors, including

- the *current state* of the movement planning field;
- *general* characteristics of the task domain, such as the presence of two bins in front of Jean;
- *specific* aspects of the current situation, such as the toy being hidden in bin A;
- *memory* of previous reaches, which bias the movement planning field in favour of previous reach directions (roughly, habit);
- competitive *interactions* between locations across the movement planning field, which help guarantee that one direction "wins out."

Change over time in these further factors is specified by their own functions, and when all these are coupled together the result is a rather complicated beast. Fortunately however the dynamical system specified by this grand equation can be simulated on a digital computer, and so the behavior of the model can be compared with the mass of experimental data gathered on the A not B error. The bottom line is this. It is possible to choose a specific set of parameters for the grand equation such that the resulting dynamical system reproduces the classic A not B error, including the various contextual subtleties that caused so much grief for earlier kinds of explanations. This in itself is an impressive achievement, since as Thelen et al. point out, there was no guarantee in advance that this would be possible. The *form* of the equations build in strong assumptions about the *general nature* of the dynamical system responsible for reaching The fact that there exists a set of parameters generating appropriate behavior within the constraints of those assumptions already goes a considerable way towared vindicating those asusmptions.

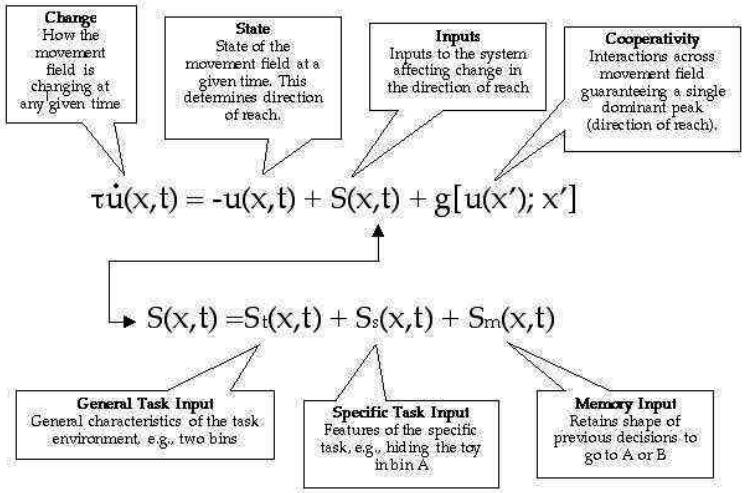

Figure 4. The "grand equation" defining Thelen et al's model of reaching in their account of the classic "A not B" error.

However, the critical test for the model is not whether it can find a way of accounting for the existing data, but whether the very *same* equations and parameters can account for any *other* relevant data that may come along. And indeed what Thelen et al found was that the model makes a number of novel predictions which were borne out in further experiments. So the exercise is not mere "curve-fitting;" it is finding a particular "curve" which not only fits existing data but successfully predicts new data.

For anyone interested, the gory details are recounted in the manuscript "The Dynamics of Embodiment" currently under consideration at *Behavioral and Brain Sciences*. My interest here is not in whether the Thelen model is, in the end, the one true account of the A not B error. Rather, I am interested in the *kind* of explanation they are providing. Thelen et al summarize it this way:

> The model accomplishes all this without invoking constructs of "object representation," or other knowledge structures. Rather, the infants' behavior of "knowing" or "not-knowing" to go to the "correct" target is emergent from the

complex and interacting processes of looking, reaching and remembering integrated within a motor decision field.

I would like to highlight just two of the fundamental differences with more traditional explanations of cognitive processes. First, the explanation looks in the first instance not at what Jean is *thinking*, but at what he is *doing*. The primary focus is not on the little Jean-like homunculus within Jean's head, but on the whole embodied Jean embedded in his environment (i.e., jeans and all!). Of course, within-the-head, neurally-instantiated processes are *involved* in the explanation of the A not B error; after all, a headless Jean wouldn't be reaching anywhere (except in a horror movie). But in the Thelen explanation, it could not be sufficient to advert to such processes; they are only one crucial part of an overall story which essentially invokes the whole embodied and embedded Jean.

Second, in this explanation there are no symbols, rules, representations, algorithms, etc., postulated in Jean's mind. Rather, the explanation is cast in terms of the continual evolution and interaction of a set of coupled continuous variables, as described by a differential equation. The A not B error is a behavior which emerges from all this ongoing interaction under certain specific conditions. If we grant that the A not B error is a genuinely cognitive issue, then we have a thoroughly dynamical explanation of a cognitive phenomenon, one in which the processes involved resemble Watt's dynamical governor much more than any orthodox computuational alternative.

4. The Dynamical Hypothesis in Cognitive Science

After these brief illustrations, it is now time to return to the main issue: what is the essence of dynamical cognitive science, and how does it differ from traditional computational cognitive science? As mentioned, my strategy in answering this question has been to note that where traditionalists have rallied under the slogan that cognitive agents are digital computers, dynamicists fall in behind the idea that cognitive agents are dynamical systems. The challenge is then to say, in a reasonably rigorous and interesting way, what this *means*. My response to that challenge, in a nutshell, is this:

The Dynamical Hypothesis (DH):

- The Nature Hypothesis: For every kind of cognitive performance exhibited by a natural cognitive agent, there is some quantitative system instantiated by the agent at the highest relevant level of causal organization, such that performances of that kind are behaviors of that system.
- The Knowledge Hypothesis: that causal organization can and should be understood by producing dynamical models, using the theoretical resources of dynamics, and adopting a broadly dynamical perspective.

OK… but what does all *that* mean? Here we have to do quite a bit of unpacking. A natural place to start is with the notion of a *dynamical system*.

Dynamical systems

Dynamical systems are, obviously, *systems* of a particular sort. A system, in the sense most useful for current purposes, is a set of variables changing interdependently over time. For example the solar system of classical mechanics is the set of positions and momentums of the sun and planets (and their moons, and the asteroids…). The question then is: when is a system *dynamical*? Interestingly, there is no established answer within cognitive science or even outside it. A search of the literature reveals a wide range of definitions of dynamical systems, ranging from the very specific ("a set of bodies behaving under the influence of forces") to the hopelessly broad ("a system which changes in time"). Somewhere on this spectrum lies the definition which is the most useful in articulating the dynamical hypothesis in cognitive science. Which is that?

The clue, I think, is found in the fact that in all those systems standardly counted as dynamical systems in the practice of cognitive science, the variables are numerical, in the sense that we can use numbers to specify their values. Why is this? Well, one thing about numerical quantities is that it makes sense to talk about *how far apart* any two values are, and indeed we have an easy way of telling what that distance is. And when a system's variables are numerical, we can also tell how far apart any two overall states of the system are. And the key point is this. For some systems, it is possible to describe how they change over time—their behavior—by specifying how *much* or *far* they change in any given time step or period. The rule capturing this description is a difference or differential equation.

In my opinion the best way to articulate the dynamical hypothesis is to take dynamical systems to systems with this property, i.e., *quantitative* systems. There are various reasons for this. First, it reflects pretty well the actual practice of cognitive scientists in classifying systems as dynamical or not, or as more or less dynamical. Second, it is cast in terms of deep, theoretically significant properties of systems. For example, a system that is quantitative in state is one whose states form a *space*, in a more than merely metaphorical sense; states are *positions* in that space, and behaviors are paths or trajectories. Thus quantitative systems support a geometric perspective on system behavior, one of the hallmarks of a dynamical orientation. Other fundamental features of dynamical systems, such as stability and attractors, also depend on distances. Third, the definition sets up a solid contrast between dynamical systems and digital computers, essential if we want to understand dynamical cognitive science as distinctively different from orthodox computational cognitive science.

OK, so we have a fix on dynamical systems; what does it mean to say that cognitive agents *are* those things? Here things will be clearer if we make a distinction between what I call the Nature and Knowledge hypotheses.

Nature Hypothesis

The nature hypothesis tells us something about reality itself, i.e., that things of one kind (cognitive agents) *are* things of another kind (dynamical systems). The truth or falsity of the nature hypothesis is completely independent of what we happen to think or know about reality; it is about the way the world is. And to say that cognitive agents *are* dynamical systems is to make a somewhat complicated claim. Notice first of all that it is not a straightforward *identification*. Jean is a cognitive agent, and one thing for sure, Jean is not simply a set of interdependent variables, just as the Watt governor is not simply the arm angle variable. The simple slogan is really saying that for any cognitive performances of mine you might be interested in, there is some set of variables associated with me (and the relevant environment) which constitute a dynamical system of a particular sort, and the cognitive performances are behaviors of that system. So for example Jean's "deciding" to reach for one box or another is a kind of cognitive performance, and Thelen et al's account suggests that associated with Jean there are various variables tied together and changing in the way specified by their grand equation, such that his reaching behavior (including his A not B error) is the behavior of that set of variables. And the nature aspect of the dynamical hypothesis says that all cognitive performances are like that. Note that on this analysis each cognitive agent "is" no *one* dynamical system; different kinds of cognitive performances would be the behavior of different systems associated with the same agent.

Knowledge Hypothesis

While the nature hypothesis is a claim about reality, the knowledge hypothesis is a claim about cognitive science. It says that cognition (at least in the case of natural cognitive agents, such as humans and other animals) is best *understood* dynamically. This is of course *because* cognitive agents are in fact dynamical systems (the nature hypothesis), and your intellectual tools really ought to fit the subject matter at hand. Conversely, our best *evidence* for the nature hypothesis would be discovering that the best way to study cognition is to use dynamics. However we should not allow the undeniable fact that the nature and knowledge hypotheses are intimately related to cloud this important distinction.

What is it to understand natural cognition dynamically? I said above that the easiest way to pick out a dynamicist in cognitive science is to see whether they use differential or difference equations, and while this is not the whole story it is certainly

a key part of it. A thoroughly dynamical perspective on cognition has three major components: a dynamical *model*, use of the intellectual tools of *dynamics*, and adopting a broadly dynamical *perspective*.

A dynamical model is an abstract dynamical (i.e., quantitative) system whose behavior is defined by the scientist's equations. The behavior of the model is compared with empirical data on the cognitive performances of human subjects. If the match is good, we infer that the cognitive performances simply are the behavior of relevently similar dynamical systems associated with the subjects. So for example Thelen et al define an abstract dynamical model by specifying a set of variables and a grand differential equation governing their interdependent change. They then show that if the parameters are set right, the model system behaves just the way Jean does; from which they infer that Jean's cognitive performances are, in reality, the behavior of a very similar system whose variables are aspects of Jean himself and his environment.

One problem with this whole approach to cognitive science is that the behavior of even simple nonlinear dynamical systems can be rather hard to understand. So while defining an abstract dynamical model might be easy enough, understanding what it does—and so whether it is a good model—can be pretty challenging. One handy tool here is the digital computer, used to simulate the dynamical model. Note that in such cases the digital computer is not *itself* a model of cognition; it is just a tool for exploring models. But much more important than the computer is the repertoire of concepts and techniques loosely gathered under the general heading of "dynamics." This includes *dynamical modelling*, that traditional branch of applied mathematics which aims to understand some natural phenomenon (e.g., the solar system) via abstract dynamical models; and also *dynamical systems theory*, the much newer branch of pure mathematics which aims to understand systems in general and nonlinear dynamical systems in particular. To use the intellectual tools of dynamics is to apply this body of theory (suitably modified and supplemented for the purposes at hand) to the study of natural cognition.

The third component of dynamical understanding is a broadly dynamical perspective. The best way to convey this somewhat nebulous idea is to describe it as the difference between the two ways of conceiving the steam governing problem. 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. Cognitive performances are conceived as continual movement in a geometric space, where the interesting structure is found *over* time rather than statically encoded *at* a time. Interaction with the world is a matter of simultaneous mutual shaping rather than occasional inputs

and outputs. Dynamicists are certainly interested in "within-the-head" structures and processes, and usually even allow that some of these count as representations, but they reject the idea that cognition is to be explained exclusively in terms of internal representations and their algorithmic transformations.

It is hard to overemphasize how different dynamical cognitive science is *in practice* from its orthodox computational counterpart, and also hard to convey the nature of the dynamical approach in a few short paragraphs. In my opinion the Dynamical Hypothesis, as formulated above, comes pretty close to encapsulating the theoretical essence of the dynamical approach; further, the contrast between the DH and the computational hypothesis is the most significant theoretical division in contemporary cognitive science. However these are are contentious *philosophical* claims about the nature of cognitive science. How have they been received by other cognitive scientists?

6. Some Objections to the DH

The DH is not true.

The largest and most considered set of responses to the DH were the set of peer commentaries in Behavioral and Brain Sciences. A rough count indicated that a majority of this self-selected bunch were basically sympathetic to the DH (in some form), and almost everyone was willing to grant that the DH (in some form) is true of at least *some* of cognition. Nevertheless one of the most common responses was to deny that the DH is true in general. This denial was grounded in the belief that at least some cognition (generally "higher" or more "central" aspects) are clearly best accounted for in computational *rather than* dynamical terms. For example Alan Bundy claimed that

> …with our current experience of the modelling power of dynamical versus symbolic techniques, this [dynamical accounts of higher level cognitive processes] seems *very* unlikely.

However such objections missed the point of my work, the whole thrust of which was to articulate the DH in order that its truth might be evaluable, rather than to argue for its truth. The difference between proposing a hypothesis for empirical evaluation, and endorsing that hypothesis as true, is a subtle one—too subtle, it seemed, for many of the commentators. My own official position is that we are not currently able to say with any certainty to what extent the DH or the competing CH are true. It will only be after lots of hard work producing and evaluating particular models of particular aspects of cognition that we will justified in asserting any verdict.

It is true that at various places I have provided broad philosophical arguments in favor of the truth of the DH, and these may have led some people to conclude that I

was already committed to DH being unqualifiedly true. However, while these philosophical arguments may be interesting but they are rarely if ever decisive. They should be interpreted, I think, not as demonstrating that the DH is *in fact true*, but as demonstrating that the DH is currently *sufficiently plausible* to be worth taking seriously, i.e., to be worth devoting the huge amounts of time and resources required for serious empirical evaluation.

Where I stand my ground is not on the blanket truth of the DH, but on the idea that the DH takes a certain form, i.e., that it should be articulated a certain way. I think these philosophical issues can be largely resolved in advance of the hard empirical work. Indeed, the corresponding philosophical questions in the case of the CH *have* been largely resolved over the past few decades. The challenge is to reach a similar level of clarity and consensus for the DH—something that my formulation of the DH has not yet achieved, to judge by most of the responses.

Eliminate the Nature Hypothesis!

The DH as I articulate it has two intimately interconnected components, one that says something about cognitive agents and one that says something about cognitive science (i.e., the Nature and Knowledge hypotheses). Another common response has been to insist that the DH is really only the Knowledge hypothesis. This idea comes in many flavors, but the thrust is to deny that dynamicists are concerned with the way the world really is. For example, the eminent dynamicist Randy Beer has argued that

> As mathematical formalisms, both computation and dynamics are sufficiently broad that there is no empirical fact of the matter about which kind of system a cognitive agent *is*…What the debate between computational and dynamical approaches to cognitive science is really about is which is the most insightful, explanatory, penetrating and parsimonious stance to take toward a cognitive agent. (Beer, 1998)

Steven Quartz claims that

> the crucial distinction between the computational and dynamical hypotheses is an epistemic one resting on the appropriate level of explanation for understanding cognitive systems. (Quartz, 1998)

and Bernstein & van de Wetering claim that

> the DH as a whole is pragmatic in nature, i.e., "it is more convenient / enlightening / interesting to describe cognition in dynamical terms than in computational terms. (Bernstein & van de Wetering, forthcoming)

Now it is an interesting thing about scientists that they are often very hesitant to use terms like "truth" and "reality," despite the fact that they more than anyone else are able to uncover the truth about reality. These scientists correctly observe that any particular scientific claim or theory may (or even probably will) eventually turn out to

be false, and that any good scientist should avoid dogmatism and acknowledge the uncertainty associated with their position. However they mistakenly go on to conclude that scientists are not (or should not be) purporting to describe reality itself, but *merely* providing more and more useful ways of talking. That is, they revert to a kind of *instrumentalism*, according to which scientific theories are only more or less convenient instruments or tools, and do not describe accurately or truly the way the world actually is. Put another way, they adopt a form of what philosophers know as Kantian transcendental realism, according to which the world "as it is in itself" is intrinsically unknowable; all we can access is the world "as understood by us."

Now this is not the place to debate the virtues or otherwise of transcendental realism. Suffice to say that for practical purposes a naïve realism is the optimal metaphysical stance. Scientists are in the business of finding out what the world is like They do so by developing successively more adequate (convenient / enlightening / interesting etc.) descriptive frameworks, where the adequacy of the framework is a matter of fit between that framework and the world. A good scientific theory is not merely useful or convenient; it asserts (correctly or incorrectly) that the world is a certain way and not some other way. So for example Thelen et al are claiming that the A not B error is the emergent behavior of a particular kind of dynamical system, and the result ill-formed concepts in the Jean's head.

This is the commonsense interpretation of what is going and we'd want pretty good arguments before surrendering it. Good arguments, however, are exactly what Beer and co. don't provide. Beer, for example, attempts to argue that the Nature hypothesis is incoherent, but his arguments turn on misunderstanding the technical details of the definitions of dynamical systems and digital computers as kinds of systems. (For elaboration, see (van Gelder, 1998a)). Bernstein & van de Wetering claim that the distinction between the Nature and Knowledge hypotheses is "unhelpful" because the Nature hypothesis doesn't *add* anything to the Knowledge hypothesis. Well, here—putting it bluntly—is what the Nature hypothesis adds:

- I have been arguing that it adds *truth*; i.e., the idea that cognitive agents *are in fact* dynamical systems, and not merely conveniently describable as such.
- When articulating the DH, distinguishing the Nature and Knowledge hypothesis enables one to sort out a whole lot of issues into two separate piles. One pile is ontological pile; it consists of issues such as: what are systems; how do systems relate to each other; what are dynamical systems; what are digital computers; how do dynamical systems and digitial computers relate; how do cognitive agents and dynamical systems relate; etc.. The other pile is epistemological; it consists of issues such as what is a model and how do models enable us to

understand natural phenomena; what is dynamical modeling; what is dynamical systems theory; what is a dynamical perspective; what are the important differences between a dynamical perspective and an orthodox computational perspective; and so forth.

Of course, in the day-to-day practice of cognitive science, there is no need to append the claim "and this theory truly describes the way cognitive agents really are" to one's dynamical theory of cognition; that much is implicit in the fact that one is asserting and defending the theory. But the *philosopher* of cognitive science would be delinquent if he *didn't* discuss such issues.

The DH is not falsifiable.

Another sort of objection is that the DH fails to be a genuine empirical hypothesis because it is not *falsifiable*, i.e., nothing could prove that the DH is wrong. One source of this objection seems to be the idea that the DH must be true of everything, and you can't falsify a theory that is trivially true. Another line of thought seems to be that the DH as formulated makes no specific empirical predictions, and so can never be tested. Let me take these in turn.

If we are fuzzy enough about what dynamical systems are and what it is to be one, then the DH certainly *would* be trivially true. However in articulating the DH I put considerable effort into crafting a hypothesis that is as narrow and precise as possible given the diversity of dynamical reseach in cognitive science. Recall that the Nature hypothesis is that claim that

> For every kind of cognitive performance exhibited by a natural cognitive agent, there is some quantitative system instantiated by the agent at the highest relevant level of causal organization, such that performances of that kind are behaviors of that system.

Note that this definition interprets dynamical systems as quantitative systems, which are a specific subclass of systems, viz., systems for which there exists metrics over their state set (and perhaps over the time sets) such that system behavior is systematically related to distances as measured by those metrics. Not every system is like this and so the Nature hypothesis is nontrivial in claiming that cognitive agents are systems of a specific kind. Second, note that the Nature hypothesis requires that cognitive agents be dynamical systems (in this sense) *at the highest relevant level of causal organization* for a given kind of behavior. Digital computers do not satisfy this condition, even though they may instantiate any number of other dynamical systems at various levels. So the Nature hypothesis requires a quite specific kind of *relationship* between cognitive agents and dynamical systems.

The non-triviality of the DH is also obvious when we consider the Knowledge hypothesis, which basically claims that cognitive agents can and should be understood in dynamical terms. If *this* were trivially true, we would already have perfect models of every aspect of cognition and cognitive science would be over. But understanding cognition dynamically is obviously *not* a trivial matter. Understanding some natural phenomenon in dynamical terms is never simple, and if anything it is especially difficult in cognitive science. After all, physicists have been producing good dynamical models since the seventeenth century; three hundred years later in cognitive science we are only just getting into the game.

In short, the DH is certainly not true of everything, and proving that it is true of cognitive agents (IF it is!) is damned hard.

The second version of the falsifiability objection is more interesting. Randy Beer suggests that the DH

> isn't a genuine scientific hypothesis, at least not in the traditional sense of making an empirically falsifiable claim. What's at issue here aren't experimentally testable predictions...

and another serious dynamicist, Richard Heath, worries that

> there is little guidance on how such investigation can determine the relative validity of DH and CH. It may be the case that it is very difficult indeed to provide the empirical evidence needed to reject CH in most cognitive scenarios, using tools available to experimental psychology. (Heath, 1998)

These are important points, and the proper response consists in explaining in what way the DH, like any hypothesis of its kind, is empirically contentful and hence falsifiable. Beer and Heath are correct the DH cannot be decisively tested by means of any direct and immediate confrontation with reality. It is a very general hypothesis, perched deep in the web of theory, and surrounded by a wide buffer of auxiliary hypotheses and chains of inference. The DH does however issue one major prediction—that our best accounts of cognition will in the long run be dynamical in form. The DH will be known false if, after an extensive period of investigation, cognitive scientists have in practice rejected dynamical approaches in favor of some other modelling framework.

In this respect, the DH is on a par with other venerable scientific doctrines. For example, the "evolutionary hypothesis," that all biological complexity is the outcome of natural selection, does not on its own make any specific testable predictions. It does however predict that in the long run all our best explanations of biological complexity will be cast in terms of natural selection. With much auxiliary theorising, the evolutionary hypothesis *does* make specific predictions, but if those predictions fail, the main hypothesis can be preserved by shifted the blame elsewhere. If there is too much

blame to be shifted, we eventually reject the main hypothesis. For broad theoretical hypotheses, this indirect connection with the world is not unfalsifiability; rather, it is what falsifiability consists in. Thus, contra Beer, the DH can be a genuine scientific hypothesis even if it alone does not make specific testable predictions.

The testability of any broad theoretical hypothesis depends essentially on a fund of good judgement which is implicit in scientific practice and can never be made fully explicit and written down in a rule book (Kuhn, 1962). Heath is right to note that in any given case it will be difficult, perhaps impossible to establish in any conclusive or mechanical way whether a dynamical model is preferable to a computational competitor, but it would be wrong to fault the DH for failing to solve this problem. Moreover, there are some very general principles that may help us even when the detailed empirical arguments are inconclusive. When scientists, as a group, choose one model or general theoretical framework over another, they inevitably allow certain very general desiderata to shape their judgements. Famously, for example, they prefer simple and elegant theories over complex and ungainly rivals; and they prefer theories which integrate well with our best theories in neighboring domains. Some refer to such virtues as “aesthetic” or “superempirical;” whatever we call them, it is clear that the process of empirical evaluation always involves relying on such criteria. This is not to say that scientific judgement is “irrational,” or just a matter of some “leap of faith”—rather, to grasp the essential role of such reliance is to understand the nature of scientific rationality.

Finally, it is worth observing that while one group of critics claim that the DH obviously *false*, another group worry that the DH is *not falsifiable*!

The truth is in the middle!

In articulating the DH, I deliberately tried to make the contrast with orthodox cognitive science as strong and clean as possible. The reason for this should be obvious enough: we are more likely to make scientific progress when the major options are clearly delineated and can stand against each other. Not surprisingly, however, some critics have claimed that the DH and CH are too extreme; that neither is likely to be true, and the truth must be somewhere in the middle. Daniel Dennett presented this idea in a memorable way. In van Gelder's view of the theoretical landscape, he claimed, there is Mt. Newton on one side, and Mt. Turing on the other, and nothing in between. The trouble is that neither classical mechanics nor Turing machines are likely to account for natural cognition. The truth about cognition will actually be found among the foothills and ranges scattered around and beyond the grand peaks.

In effect, Dennett is claiming that the DH and the CH caricature the available options in contemporary cognitive science. However Dennett only makes his point by

himself egregiously caricaturing the DH. Dynamical cognitive science is not simply (indeed, not ever) the straightfoward application of Newtonian mechanics to natural cognitive processes. The dynamical umbrella covers a rich tapestry of models and theoretical machinery, including, I think, much of the supposed middle ground between Newton and Turing.

To see this, consider first the general notion of computation. What makes a process a *computational* process? In my opinion the answer is that a computational process is one that sets up a mapping between two domains. Metaphorically, computational processes systematically provide *answers* to *questions*: provide a question as input, and the process will deliver an answer as output. In this sense, almost anything can be construed as a computer. The concept of computation only starts to get interesting when significant further constraints on the nature of the process involved. The most familiar approach is to require that the process be *effective*: intuitively, to produce its answers by means of a finite number of discrete operations specified by some finite recipe or algorithm. Effective computation is the same thing as *Turing* compution, which is equivalent to *digital* computation.

The second half of the twentieth century has come to be dominated by the digital computer. This is obviously true in practical domains, but it is also true in the intellectual sphere. For example that body of mathematics going under the name of "theory of computation" has been overwhelmingly the theory of digital computation. Closer to home, cognitive science has been dominated by the idea that natural cognition is a form of digital computation. This is the essence of orthodox approaches. Given these developments, many people seem to have lost sight of the fact that there are many other kinds of computation. "Non-Turing" computation is simply any kind of computation which for whatever reason fails to satisfy the full set of strict conditions for counting as digital computation. Thus, in the days before digital computers became widely available, analog machines such as differential analyzers and even the humble slide rule were used for everyday computational tasks. Sometime back in the 1960s *Scientific American* carried an article describing how to build your own personal computer at home— analog, of course.

Given the vast range of possible forms of non-Turing computation, it makes no sense to ask how non-Turing computation "in general" compares with its digital counterpart. But one can focus on specific *kinds* of non-Turing computation, defined by alternative sets of constraints. One approach is to consider a given class of dynamical systems as computers. There is now a whole branch of the theory of computation vigorously enquiring into the computational properties of dynamical systems performing one form or another of non-Turing computation. The existence of a rigorous

body of knowledge at the intersection of dynamical systems theory and the theory of computation obviously opens up a whole new set of possibilities for understanding natural cognition. It may well be that certain aspects of cognition are best understood as the behavior of dynamical systems performing non-Turing computation (Garson, 1996)—that is, as occupying the "middle ground" between Mt. Newton and Mt. Turing. To the extent that this is true, the orthodox computational theory of mind clearly stands refuted. Would it likewise refute the DH—or vindicate it?

Nothing in the DH requires that natural cognition be understood in *solely* traditional dynamical terms. Indeed, such a requirement would be quite bizarre. Do orthodox models draw *solely* upon the theory of computation? Dennett caricatures the DH by placing it atop Mt. Newton. In reality dynamicists draw on a wide range of auxiliary concepts, methods, etc., even while holding to their dynamical core. One strategy is to combine dynamics with non-Turing computation—to see a cognitive process as simultaneously the behavior of a dynamical system and as a kind of analog computation. This middle ground, I believe, really belongs to the dynamical approach to cognition, just in case a thoroughly dynamical perspective continues to be essential to understanding the process. If the dynamics eventually drops out and the process is understood primarily as computation—even non-Turing computation—then the DH ceases to be true of that process, even if at some level the the process is in fact the behavior of some dynamical system.

7. Conclusion

I conclude that the DH still stands as the proper way to articulate the essence of contemporary dynamical approaches to cognition. But what about the question I keep deferring: is it actually *true*? To answer this is, in effect, to predict the course of cognitive science; and, as a pundit once pointed out, it is hard to make predictions, especially about the future. Moreover, a philosopher somewhat removed from the front lines I have certainly have no special insight. However, putting qualifications aside, recent broad trends in cognitive science, as well as some very general considerations, indicate that the Dynamical Hypothesis will turn out to be true of a considerable portion of natural cognition; that where computation is relevant, it will be analog computation implemented in dynamical systems; and insofar as the DH is false, it will be superseded by some form of theoretical framework whose elements are being pieced together by unheard-of mathematicians laboring under the illusion that their ideas couldn't possibly have any application to reality.

References

Beer, R. (1998) Framing the debate between computational and dynamical approaches to cognition. *Behavioral and Brain Sciences*, **21**, 630.

Bernstein, D., & van de Wetering, S. (forthcoming) More boulders of confusion. *Behavioral and Brain Sciences*, **21**.

Dennett, D. (1995) *Darwin's Dangerous Idea*. New York: Touchstone.

Garson, J. (1996) Cognition poised at the edge of chaos: A complex alternative to a symbolic mind. *Philosophical Psychology*, **9**, 301-321.

Heath, R. (1998) Cognitive dynamics: a psychological perspective. *Behavioral and Brain Sciences*, **21**, 642.

Kuhn, D. (1962) *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Newell, A., & Simon, H. (1976) Computer science as empirical enquiry: Symbols and search. *Communications of the Association for Computing Machinery*, **19**, 113-126.

Quartz, S. (1998) Distinguishing between the computational and dynamical hypotheses: What difference makes the difference? *Behavioral and Brain Sciences*, **21**, 649-650.

van Gelder, T. J. (1998a) Disentangling dynamics, computation, and cognition. *Behavioral and Brain Sciences*, **21**, 40-7.

van Gelder, T. J. (1998b) The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, **21**, 1-14.