Real Patterns and Distributed Cognition[1]

Andrew Dellis and
David Spurrett

Introduction

Cognitive science has always been multi-disciplinary[2], growing as it did out of a recognition that cognition as a general object of study demanded expertise from a variety of fields including psychology, linguistics, computer science and philosophy. More recently, of course, biology has come to play an increasingly important role in this disciplinary mixture[3]. For all the

---

[1] This paper grew out of research conducted by Dellis during the first half of 2001, and by Spurrett in 2000 and 2001. Parts of Spurrett's contribution here were presented at the April 2001 conference in philosophy of science at the Inter University Centre, Dubrovnik, Croatia, his attendance at which was partly funded by the NRF, whose assistance is hereby acknowledged.

[2] Bechtel, Abrahamsen and Graham open their account of 'The Life of Cognitive Science' in Bechtel and Graham (1998) with the following definition: 'Cognitive science is the multidisciplinary scientific study of cognition and its role in intelligent agency. It examines what cognition is, what it does, and how it works. We don't entirely agree with this definition (see the following note) but do endorse the prominence given to the multidisciplinary character of the field.

[3] In the account by Bechtel, Abrahamsen and Graham (1998:94) relatively little emphasis is given to biology, which appears in none of their diagrams of the various disciplinary contributions to the field. Nonetheless the extensive references to biological studies in, *inter alia*, works by Brooks (e.g. 1991a; 1991b; 1997), Clark (especially 1997), and Thelen and Smith (1994) speak for themselves.

advantages brought by such a melting pot, some issues and questions can fall between the interdisciplinary cracks, and it is one of these that we want to consider here. Working scientists are not typically concerned with metaphysical questions[4], but their efforts nonetheless often raise such questions, or suggest new ways of their being framed. Metaphysicians on the other hand are not always as attentive to current science as would be desirable. The particular issue which concerns us lies at one of these intersections of empirical and metaphysical questions, and draws chiefly on two lines of thinking.

The first is Dennett's (1991) most metaphysical piece of work to date, 'Real Patterns'[5]. This paper is, in part, Dennett's response to demands that he take some definite position with respect to the question whether he is a realist or an instrumentalist about descriptions generated from his 'intentional stance'. His comments on the reality of patterns are supposed to meet this challenge. A recent critical engagement with that paper by Ross (2000) argues that the Dennettian position on a range of issues is best cashed out by taking the argument of 'Real Patterns' (with significant modifications) as precisely a statement of fundamental ontology, that good Dennettians should hold the world to be 'patterns all the way down' (Ross 2000:160).

The second line of thinking is a particular body of research pointing towards the view that cognition is, and that it *can* be, 'distributed' in the sense of exploiting non-neural resources[6]. In a series of papers Kirsh and

[4] There are cases where such questions become central, as has been persuasively argued by, for example. Koyré (1957) and others in connection with the astronomical revolution in the sixteenth and seventeenth centuries.

[5] 'Real Patterns' is reprinted in Dennett (1998). All page references to 'Real Patterns' in this paper are to the pagination of the 1998 version.

[6] There is another sense of 'distributed' typically associated with connectionist architectures, where calling a computational process distributed indicates that the ways in which information is processed by the system in question are unlike those in classical computational systems. For more detail and applications see Rumelhart and MacLelland (1986), and for an excellent overview and introduction see Clark (1989). We take it for granted in the present paper that cognitive processing in brains is distributed in broadly connectionist ways.

Maglio (1991; 1992; 1994; 1997)[7] studied the behaviour of human players of the computer game Tetris and compared it with predictions based on a classical and non-distributed model of optimal play. On the basis of differences between observed human play and the classical predictions, and supported by a number of supplementary experiments they make a compelling case for the thesis that human players exploit a variety of non-neural resources in the course of play, and coin the term 'epistemic action' to name the category of physical actions with cognitive benefits in question.

A short way of putting the question which concerns us, then, is to say that we want to know whether the structures identified by Kirsh and Maglio are real patterns in Dennett's sense. Before explaining in more detail how we see the issues, and how we want to approach that question, though, it will be worth saying a few things about why it is important.

One reason is, bluntly, to 'keep the riff-raff out'. The hard-won recognition that cognition is distributed beyond the brain and into the body and world is all too amenable to appropriation by vague, nebulous and unscientific forms of thinking. Part of the way to deal with this, in the interests of a general commitment to naturalism for which we won't argue on this occasion, is to get the metaphysics right. Feyerabend (1987; 1993), who had few metaphysical commitments but vociferously defended the perceived rights of what he called 'traditions' (which could include metaphysical positions) notoriously sought to defend astrology from scientific attack (see especially Feyerabend 1978:91-96)[8]. More recently Dupré (1993:10,263) who is in some ways sympathetic to Feyerabend has found himself struggling to come up with a good reason (he does admit to having

[7] Although see also Maglio *et al.* (1999) which extends the research to consider the game of Scrabble.

[8] A supporter of Feyerabend might object that his major purpose was to poke holes in the scientistic arrogance visible in the 1975 anti-astrology 'Statement of 186 leading scientists' which appeared in *The Humanist*. Quite so, but in the course of doing so he does draw on a range of empirical evidence for the effects of solar and other celestial activity on terrestrial life as though such evidence supports astrology *per se*. In later life Feyerabend did, admittedly, express dissatisfaction with his 1978 book (see Feyerabend 1995:147).

'prejudices') for rejecting astrology, a fact we would argue is largely explicable by reference to his commitment to a highly disordered metaphysics. Talk of mind, or at least cognition, being spread wider than the limits of the brain can too easily sound close to, for example, Huxley's (1954) peyote-induced talk of 'Mind at Large'. If this seems implausible or even alarmist, consider *The Embodied Mind* by Varela, Thompson, and Rosch (1991) who take their often subtle and interesting reflections on embodied, embedded agents, anchored in some good empirical work, to count as evidence against realist and objectivist views of body mind and world. Instead they favour a focus on Buddhist 'transformative analysis' and meditative practises in order to account for 'the basic circularly of our condition'. Clark (1997:173) expresses related worries about their programme.

A more serious reason, besides crowd control, is that current thinking in cognitive science is relevant to a variety of philosophical debates concerning mental causation and explanation but stands cut off from such debates partly for the lack of much work linking the results of cognitive science with the metaphysical concerns of philosophers. One striking symptom of this is the continued reliance in most of the mental causation literature (for a recent example see Kim 1998) on a basically Humean conception of causation, when more and more of cognitive science moves towards a dynamical systems approach. Without some bridges being built between the two areas, parts of philosophy of mind are going to remain sadly out of touch. Rather than build a bridge from scratch we hope to approach our question largely by moving an available resource into position, that resource being Van Gelder's useful paper on 'The dynamical hypothesis in cognitive science' (1998). In that paper he distinguishes between a *nature hypothesis*, to the effect that cognitive agents actually *are* dynamical systems, and a *knowledge hypothesis*, which is the more cautious, or ontologically agnostic thesis that cognitive agents can be understood dynamically. His discussions will help us get to a position where we can make some headway with what was noted above as our central question: whether the structures identified by Kirsh and Maglio are real patterns in Dennett's sense.

Although Kirsh and Maglio's research is surveyed below, we should say now that we take it as uncontentious that they have pretty much established a distributed version of Van Gelder's knowledge hypothesis with reference to human Tetris playing. But that means that they have established that such players *can* be well understood as dynamical systems which include a range of components or interacting parts, only some of them neural. If there is a case that the nature hypothesis follows, then we have the result that human Tetris players actually *are* distributed systems, that the computational problems demanded by the game and solved by the players are not handled by the brain alone. Specifically, then, the subject of the paper is the *ontology* of the cognitive agents posited by accounting for the mind as dynamically realised by both body and world. The question whether the nature hypothesis follows is one which connects up directly with the debate concerning realist and instrumentalist readings of Dennett. We think that we can make a case for the conclusion that the dynamical structures identified by research like that of Kirsh and Maglio are indeed real patterns in Dennett's sense (as refined and debugged by Ross) or at least are candidates for being real patterns.

In what follows we proceed as follows: First we set out the outlines of Dennett's argument in 'Real Patterns' and explain a little of the context in which it occurs. Then we briefly survey the accumulating evidence for distributed cognition against the backdrop of more traditional expectations. With both of those topics set up we turn to a more detailed examination of Kirsh and Maglio's research on Tetris playing. This in turn allows us to begin dealing with our central question, which occupies the remainder of this paper, beginning with a discussion of Van Gelder on the dynamical hypothesis in cognitive science.

Dennett on Real Patterns

Dennett defends three different 'stances' as ways of looking at, explaining, and sometimes predicting the behaviour of part of the world, each of which serves particular purposes and has particular limitations. The three are the *physical*, the *design* and the *intentional* stances. Being a type of physicalist (although he generally uses the label 'materialist') Dennett thinks that the physical world is causally closed, which means that any physical state of affairs can 'in principle' be causally accounted for by reference only to physical states and processes. When we take the *physical* stance to something we are concerned with the physical processes which are involved in it, and assuming that we had accurate enough measurements, physical understanding, and computing resources (the proverbial large enough piece of paper and sharp enough pencil) we could, again 'in principle', follow every detail of what happened and even make the best possible, although not necessarily deterministic[9], predictions of what would happen next.

In all except very rare cases we are not in a position to take a physical stance to any system in any detail. Furthermore, and Dennett is especially alive to this next point, what matters to us often doesn't depend on specific facts at the physical level. I just don't care what kind of polymer my drinking straw is made of, as long as it has two, and only two, holes and doesn't make my drink taste odd. This makes sense because often we're concerned with *functional* aspects of the large-scale activity of objects or systems. This is where the other two stances come in.

When we take the *design* stance towards some system we 'assume' (not necessarily as a deliberate decision) that it serves some function, and ask what that function might be, or whether it is working properly, and so forth, but largely ignoring what would be paramount from the physical stance (see Dennett 1995:229f). Design stance descriptions are *simpler* than physical ones, which is to say that they are often simple enough for us to make and understand them at all, which is almost never the case with the physical stance, but this tractability is bought at the expense of leaving out a lot of detail. This makes design stance descriptions and explanations somewhat vulnerable, but they can be very effective—with almost no physical stance information at all, I can predict pretty confidently that a drinking straw with, say, seven holes of the same size won't be much use for drinking.

Finally we can, in some cases, take the *intentional* stance, which means treating a system as having goals and some measure of rationality.

---

[9] Papineau (1993:16) puts it as follows: 'I take it that physics, unlike the other special sciences, is *complete*, in the sense that all physical events are determined, or have their chances determined, by prior *physical* events according to *physical* laws. In other words we need never look beyond the realm of the physical in order to identify a set of antecedents which fixes the chances of every physical occurrence'.

When we do this we might speak in terms of 'trying' to do *x*, or planning to do *y* or 'wanting' to achieve *z*, which is to say that we attribute beliefs and desires, or something like them, the apparatus of 'folk psychology', to the system (see, *inter alia*, Dennett 1987; 1998). For Dennett intentional stance descriptions are always interpretations, and, furthermore, are always underdetermined. Nonetheless he thinks that they are the best available strategy for dealing with some types of system in the world, most notably one another. (Just try to describe anything that you or someone else is *doing* without referring to intentions, desires, beliefs and so forth.)

It is probably not surprising, given that Dennett thinks there is a physical stance to be taken, even if only 'in principle' that there has been some heated debate over the status of what is said from the other two stances, especially the intentional. After all, most of the contemporary mental causation debates concerns problems (real or not) which arise because thinking that physics *is* causally closed seems to make any non-physical causal claim, especially those referring to beliefs and desires, seem threatened with redundancy. Dennett sometimes makes Quinean noises here, suggesting that intentional descriptions (referring to 'propositional attitudes') are a kind of *façon de parler* or 'dramatic idiom' (e.g. Dennett 1987:110), which doesn't refer in the same way that fundamental physical descriptions do, but pay their way by being *useful*. But he also goes to some lengths to make clear that there are *true* explanations which are possible *only* from the intentional or design stances. So is he a realist or an instrumentalist?

He has said that he aims to place his view 'firmly on the knife-edge between the intolerable extremes of simple realism and simple relativism' but admits that 'this has not been recognized as a stable and attractive option by many others in the field' and that his 'critics have persistently attempted to show that my position tumbles into one abyss or the other' (Dennett 1987: 97). He also insists that his view should be seen as 'a *sort* of realism, since I maintain that that patterns [visible from the intentional stance] are really, objectively there to be noticed or overlooked' (Dennett 1987:97).

Part of the problem for Dennett here is that the adoption of the intentional stance seems to involve postulating entities and processes which are both abstract and observer dependent, which encourages an instrumentalist reading. On the other hand the facts about the success or failure of the adoption of the stance, are to Dennett perfectly objective, so that beliefs and desires look likely in some sense to be real. In *The Intentional Stance* Dennett acknowledges the problem, and reiterates his choice of the knife-edge position:

> My thesis will be that while belief is a perfectly objective phenomenon (that apparently makes me a realist), it can be discerned only from the point of view of one who adopts a certain predictive strategy, and its existence can be confirmed only by an assessment of the success of that strategy (that apparently makes me an instrumentalist) (Dennett 1987:15).

There are two main components to Dennett's ongoing development of his position here: one is the defence of the *usefulness* of the intentional stance, and the other concerns the status of the 'patterns' referred to in the passages quoted above. It is the second component which is of interest to us, and we take it as more than established that the intentional stance enables the making of powerful predictions with greater efficiency than *any* other available method.

In pursuit of both projects Dennett (1987: 25-28) offers a thought experiment which for our purposes we will simplify and shorten. Suppose some Martians with superior intelligence were to land on earth. Suppose, further, they did not need the intentional stance or even the design stance to predict our behaviour in all its detail, but that instead they can comprehend the behaviour of people at the microphysical level in order to predict behaviour down to for example the next bat of you, the reader's, eyelid. Essentially, we are to imagine, they can predict the individual behaviours of all the bodies they observe without ever treating any of them as intentional systems. (We are not to suppose that they do this with themselves or each other.)

From the Martian point of view, then, we really are not believers at all. If so, it might seem, our status as believers is not objective, but is instead an anthropomorphic notion, of utility given to our limited ability to track the world at the physical level. Dennett suggests, however, that even though these Martians may be able to predict the future of the human race at a purely physical stance level, if they did not also see us as intentional systems, they would be missing something *perfectly objective*. Namely the *patterns in human behaviour* that are describable from the intentional stance, and *only from that stance*. The patterns that is, that support *our own* generalisations and predictions.

In this respect suppose that a Martian engages in a prediction contest with an earthling. The Martian, doing all his microphysical calculations, predicts the behaviour of an individual. Given that the earthling would be equally able to predict the behaviour of that individual, say a batsman after being conspicuously clean bowled, without access to the physical stance, the Martian would be left in amazement at how this was done. And it *is* amazing to think that an intentional stance prediction could sometimes, indeed often, do just as well as one generated from the physical stance, even assuming hugely unrealistic resources for pursuing the physical stance.

We know that we can make generally effective predictions of the behaviour of intentional systems without paying attention to their microphysical constituents. The question, though, is *why* we are able to do this, or what we are latching onto when we do. Dennett's answer is that there are high level *patterns* in the activity of intentional systems, and that we can and do track these patterns.

In fact, Dennett is committed to their being a range of 'higher level' patterns, not only ones associated with intentional systems. He illustrates the key idea by means of a discussion of Conway's Life, the cellular automata designed to test some of von Neumann's ideas concerning replication[10]. The universe of Life is a two-dimensional grid, the cells of which can be either occupied or unoccupied. Time in Life moves discretely from generation to generation, and the state of the next generation is determined by the following rules, applied to each cell in the grid: If the number of occupied neighbours of a cell is *two*, that cell stays in its current state into the next generation, if the number of occupied neighbour cells is *three*, the cell will be occupied in the next generation, and in all other cases the cell will be unoccupied. That's all simple enough, what is interesting is what can come out of such a system of rules.

In an important paper on general features of cellular automata Wolfram classifies cellular automata into types based on their dynamical

[10] Dennett (1995) also discusses *Life* in related ways.

properties. He divides them into four types (Wolfram 1984:5), those, class I, which converge quickly onto and remain in a homogenous state, those, class II, which result in separated simple configurations of occupied cells and/or converge on a periodic pattern of states, those, class III, which result in some or other chaotic pattern, and finally those of class IV, which produce 'complex localized structures, sometimes long-lived.' These options correspond roughly to standard dynamical possibilities, in that the first class amounts to a system with a point attractor, the second to a periodic attractor or limit cycle, and the last two to 'strange attractors'[11], or at least to configurations and sets of rules which take a very large number of generations to converge on some more stable state of class I or II. Classes III and IV are of most interest, since they manifest the most interesting structures and also have the property that prediction of what structures will result from the application of the rules to a given starting situation is not typically determinable by any means except direct simulation of the rules and configuration (Wolfram 1984:31; see also Wolfram 1983).

The structures possible in Life are rich and varied, including 'blinkers', 'gliders', 'spaceships', 'puffer trains', 'glider eaters' and so forth. A glider is most easily described as a shape which moves through Life-space, contorting through a fixed sequence of transitional states like a little digital caterpillar. Watching Life run on a computer screen and seeing a glider it is all but impossible to think of the glider as a moving thing, but there's the rub: from the bottom level of description there are 'only' cells which are either on or off, and no motion takes place. Are gliders, then, and their motion, or any of the other furniture and behaviour of the Life world, real?

Dennett points out that it has been established in principle that a Universal Turing Machine could be constructed on the Life plane, and invites us to imagine one which also happened to be running a chess playing program. The required array would be huge, requiring about 10^{13} pixels according to some estimates (Dennett 1998:109) A range of levels of description of such a Life array are possible—from one concerning single

[11] Lorenz (1963) is the classic reference here, although he himself does not use the term 'strange attractor' in this paper, referring to the time development of his artificial weather as 'quasi-periodic' instead.

cells, through ones concerning gliders and other objects, right up to ones interpreting macro-states of the array as chess moves. The 'higher' the level of interpretation, the less work is needed to make predictions, including good ones such as 'the next move should be an attempt to save the Queen, probably QxP'. Indeed such predictions can be routine and relatively easy, whereas applying the transition rules of Life to trillions of individual cells is prohibitively demanding.

The 'objects' to which we refer when making these higher level descriptions, interpretations and predictions are abstract objects, according to Dennett, what Reichenbach called *abstracta*, the same type of objects as centres of gravity, or 'Dennett's lost sock centre: the point defined as the center of the smallest sphere than can be inscribed around all the socks [Dennet has ever] lost in [his] life' (Dennett 1998:97). Reichenbach's *abstracta* were to be contrasted with *illiata*—the more fundamental level 'physical' properties of something, which in the Life case would be the states of individual cells. Dennett, arguing that metaphysically speaking all abstracta are equal, says we should concentrate on which ones pay their way by giving us explanatory, descriptive or predictive leverage, which he thinks intentional stance descriptions do exceptionally well. The 'good' abstract objects, he says, are those which 'deserve to be taken seriously, learned about, used' (Dennett 1998:97).

While partly dodging the demand for an account of reality which would answer the question whether he is realist or instrumentalist about even 'good' abstracta, Dennett also proposes that a kind of realism about patterns is plausible, actually 'demonstrably right' (Dennett 1998:97), drawing on the information theoretic notion of compressibility: patterns, unlike randomness, are compressible. Patterns, in this sense, 'are there to be described—whether or not we care to see them' (Dennett 1987:28). And, we might add 'whether or not we care to get metaphysical about them.' Before we can start asking our question about the status of the patterns discovered in Kirsh and Maglio's research, though, we need to say a little about distributed cognition.

Distributed Cognition

It can seem natural and innocuous to suppose that cognition, or thinking, is something done exclusively by the brain, or at least the central nervous system. For some time now the dominant image of cognition has been that it is *some* kind of computation. While controversy and debate continue on exactly what kind of computation it might be, there is something like a consensus that some kind of information processing will be going on. Not only that, brains look like splendid candidates for organs which do the processing, especially in the light of the past few decades of work on neural networks and ever finer understanding of the functioning of the brain itself. Saying that brains are (*inter alia*) powerful information processing engines of some sort, though, is not the same as saying that all information processing goes on in brains. These two distinct claims can all too easily be elided, partly because it can seem intrinsically implausible that something which doesn't have a relatively conspicuous information processing architecture, which a huge collection of connected neurones clearly *does*, could be doing any computing at all. With Cowley and Spurrett (forthcoming) we call the view that all cognition does take place in the brain 'cognitive internalism'.

There is mounting evidence, though, that cognitive internalism is misguided, and that human and animal cognition, not to mention the most effective and flexible artificial cognition, exploits non-neural resources, i.e. bodily and/or environmental ones, in a wide range of ways. An excellent survey of research of this type is Clark (1997), although see also Thelen and Smith (1994), Brooks (1991a). Opposed to cognitive internalism, then, is the view that cognition not only can be, but typically is distributed. The notion of solving problems by leaning on the environment is, of course, not new. Vygotsky's (1968) 'zone of proximal development' concerns what the individual may achieve with the aid of an external guide. But recent work on distributed cognition tends to find that distributed cognition is not a developmental phase, but an ongoing feature of effective cognition. This work also concerns itself with developing finely structured models of the distributed processes which it identifies and studies[12]. For our purposes, as noted above, we will focus on a particular example, which is Kirsh and Maglio's research on Tetris-playing behaviour.

[12] For a brief survey of the case for distributed cognition see Spurrett (this volume).

Kirsh and Maglio on Tetris Playing

Tetris is an excellent arena for testing hypotheses about cognition. It is a game played at a computer, in which players attempt to position various falling shapes (which Kirsh and Maglio call 'zoids') so that they completely fill horizontal rows across the bottom of the field of play, and result in the player being awarded a number of points. Completed rows disappear, and allow whatever is above them to move that number of rows downwards as a result. Incomplete rows become buried under poorly placed zoids, which reduces the amount of space, and hence time, available for placing new pieces. The game ends when the field of play becomes so clogged with incomplete rows that no new zoids can descend from the top. Besides the way in which misplaced zoids increase difficulty, the game automatically increases the rate with which zoids fall as a function of accumulated points. The player task of matching zoid shapes into the contour formed by zoids already in place, is effected by four possible actions that the player can perform with single keystrokes. They are: (1) rotate a zoid 90 degrees contraclockwise, (2) translate it one step to the left, (3) translate to the right, and (4) drop. The latter involves moving the zoid instantly from its current position to the position it would eventually come to rest if no more keys where pressed. Only one zoid is ever in motion at a time, and a new one randomly selected from the available types appears the moment the preceding one comes to rest, or is dropped (Kirsh & Maglio 1994).

The game, then, has definite objectives, and unambiguous criteria of success and failure. It makes increasing demands which are highly time-dependant, and it is possible to gather a range of temporally precise data about player behaviour some of which can be easily interpreted and evaluated given that fact the aims of the game are specified in advance[13]. As noted in Spurrett (this volume), Kirsh and Maglio were led by their investigations to propose a category of 'epistemic' actions, to be contrasted

[13] Similar advantages follow from Hutchins' (1995) decision to study co-operative cognition in teams by focussing on navigation, which yields a similarly quantitative backdrop helping interpret and evaluate the activities of navigation teams.

with 'pragmatic' actions[14]. A pragmatic action is one undertaken to bring an agent closer to some physical goal, whereas an epistemic action (also a physical action) is performed in order to modify the computational state of the system performing the action. A more detailed account of epistemic actions states that an epistemic action is one which improves cognitive performance by:

1. reducing the memory involved in mental computation, i.e., space complexity;
2. reducing the number of steps involved in mental computation, i.e., time complexity;
3. reducing the probability of error in mental computation, i.e. unreliability (Kirsh & Maglio 1994).

Kirsh and Maglio's empirical case for the occurrence of epistemic actions with respect to Tetris playing is founded on a contrast with what a classical cognitive internalist approach to the game would be. They develop a model of how such an approach would deal with the computational problems posed by the game, which breaks processing down into the following four stages:

1. Create an early, bitmap of representation of selected features of the current situation;
2. Encode the bitmap representation in a more compact, chunked, symbolic representation;
3. Compute the best place to put the zoid;
4. Compute the trajectory of moves to achieve the goal placement (Kirsh & Maglio 1994).

Using this model Kirsh and Maglio predicted the number of zoid rotations which would be optimal, and then observed the actual behaviour of human players. The model predicts that rotations should only take place *after* the optimal zoid placement has been established, which is not what happens at all—rotations and translations are far more abundant than the traditional

[14] Kirsh and Maglio (1992) describe an additional category of 'perceptive' actions.

view predicts, and they start earlier than it can account for. In fact, they often start *before* a new zoid is completely visible, which in terms of the classical model would imply that they were entirely unmotivated.

Of course Kirsh and Maglio can't, and don't, just say that these actions have to be 'epistemic' actions. Rather, by means of a series of subsidiary experiments, they show how the efficiency of human computation is improved by means of them, in a range of ways. These include demonstrations that rotated zoids can be more quickly identified; that engaging in physical rotations simplifies the problem of detecting 'fit' between a new zoid and the contour below; and that performing translations which 'bounce' zoids off the walls of the playing area reduce the likelihood of vertical alignment errors. Finding the rationale, or the pay-off, for engaging in these actions, though, means abandoning not only the details but the foundations of the traditional view Kirsh and Maglio used to generate predictions and empirical contrasts. Computation, or Tetris cognition, isn't all 'in the head' but rather draws on a range of resources extending beyond the brain, which are closely coupled together to achieve optimal overall performance.

For our purposes it is important to note that one result of Kirsh and Maglio's work is a kind of specification of the high level, or functional, properties of a distributed cognitive architecture. The cognitive system they discovered and partly analysed is at some remove from a low level neural and muscular description, and vastly removed from a bottom level 'physical stance' description in terms of fundamental particles. Rather their analysis highlights, partly by means of a kind of 'reverse engineering'[15] what it is that brain, body and external systems (especially screen and keyboard) contribute to dealing with the computational demands of Tetris, and why these resources do a better job in just *those* combinations rather than others. Their work is not concerned with *how* exactly fingers, brains, eyes, screens and so forth do what they do, and so components which were functionally equivalent but structurally different would in principle be interchangeable would in a sense be indistinguishable to their enquiry. What they find and describe is at the level of the design stance, given that an overall intentional

[15] Dennett (1995:212-219) discusses reverse engineering and 'artefact hermeneutics' *inter alia*.

stance reading comes for 'free' with the rules and objectives peculiar to the game of Tetris.

Let's call the architecture that they discover a 'structure' to leave open whether it is a 'pattern' or not. As a structure, it is a specification of a distributed, or coupled cognitive architecture. Now we ask whether it is the type of thing a Dennettian should want to be a realist about. As we said above, we use Van Gelder on dynamical systems as something of a bridge between empirical and metaphysical questions, and having asked about the reality of Kirsh and Maglio's structure, we leave that question hanging for a while to take a look at Van Gelder's views. Although Van Gelder refers to a range of examples of empirical cognition research, Kirsh and Maglio's work is not on his list of examples.

Van Gelder on Dynamical Systems

In his 1998 publication, Van Gelder sets out to clarify and defend what he calls the 'dynamical hypothesis' in cognitive science, which is the hypothesis that cognitive agents are dynamical systems. He argues that this view could replace[16] what he takes to be the still-dominant computational hypothesis, holding that cognitive agents are digital computers. Van Gelder's contention is that digital computers and dynamical systems are importantly different types of system, and the ways in which he makes and defends this claim are important for our own argument here.

For his purposes Van Gelder takes a system to be a 'set of interdependent variables' where the state of the system is 'simply the state or value of all its variables at a time' and the behaviour of the system 'consists of transitions between states' presumably depending on the precise nature of the interdependence between the variables (Van Gelder 1998:616). On this wide understanding it is clear that both digital computers and dynamical systems are going to be examples of systems. Before turning to the problem of distinguishing the two, which he clearly must if his thesis is to make

[16] Van Gelder (1998:627) takes pains to indicate that whether the dynamical hypothesis will replace the computational hypothesis would depend on the outcome of 'sustained empirical investigation'.

sense, Van Gelder offers one further distinction, which is that between 'concrete' and 'abstract' systems.

A 'concrete' system is, for Van Gelder, a system such as the solar system in which the variables in the system (he is presumably referring to masses, positions, momenta and so forth) 'are actual features of the real world changing in real time in accordance with natural laws' (Van Gelder 1998:617). Although he makes no mention of Dennett (who does not appear at all in his list of references), note the similarity between a concrete system in this sense and the type of consideration which would be relevant when taking the physical stance. An 'abstract' system, on the other hand, is a set of 'abstract variables governed by mathematical rules' (Van Gelder 1998:617). Sometimes we can build or find a concrete system which 'realises' an abstract system, Van Gelder's example being a physical calculator which realises a given 'abstract computational system'. In addition to this, Van Gelder sets up what he calls a relationship of 'implementation' in which a low level system implements a higher level system only when the variables of the higher level system are 'somehow constructed' out of variables of the lower level one. The implementation relationship is weaker than full identity, but does allow some degree of identification at the level of behaviour.

Van Gelder considers the most significant features of digital computers to be that variables and states of such systems are themselves digital, that time is discrete, state changes or computation depend on an algorithm, and that the states and behaviours of the system admit of systematic interpretation (consider the calculator example above). The level at which we are typically interested in a digital computer is that governed by the interpretation, rather than with details of the concrete system itself, on which the object of our interest is implemented.

With respect to dynamical systems, on the other hand, Van Gelder emphasises the importance of the fact that variables or states in such systems are typically quantitative rather than discrete, that time does not typically operate discretely and further that quantitative states can depend on time, and that rates of change can play a significant role in fixing future states. Van Gelder proposes to take a system as dynamical 'to the extent that it is quantitative in one of the above senses' (Van Gelder 1998:619). An obvious interjection at this point would note that on the basis of the characterisations just set out digital computers simply *are* dynamical systems, which looks *prima facie* awkward for Van Gelder's main project. This is, of course, entirely true—digital computers are at basis concrete systems in Van Gelder's sense, and they undergo changes which means that *some* dynamics must be true of them[17]. Van Gelder is entirely aware of this, and also of the fact that many of the dynamical systems he has in mind are studied by means of software which is run on the very digital computers which he thinks are a bad model of cognition. Before looking at how Van Gelder deals with the objection it is worth pausing to note a similarity with the case of Dennett. Recall the complaint that Dennett's account of intentional stance descriptions is supposedly doomed to collapse into instrumentalism just because the alternative physical stance description seems to be holding the causal trumps. Dennett's position is supposed to show how it can be OK to be a kind of realist about intentional descriptions without weakening his commitment to the physical stance. One can readily imagine a committed 'digitalist' saying that Van Gelder might well be correct to say that interesting and important things can be said about dynamical systems, but that at bottom we know about them because of things we have done with digital computers, and hence that digital approaches are still the ultimate causal and explanatory winners[18].

As noted above Van Gelder distinguishes between a 'nature' and a 'knowledge' hypothesis, each of which he takes to be a component of the dynamical hypothesis. The nature hypothesis states that cognitive agents *are* dynamical systems, while the knowledge hypothesis states that cognitive science should 'take dynamical form', i.e. that cognitive agents are best studied as dynamical systems. The objection presently on the table effectively grants the knowledge hypothesis but refuses to allow an inference from it to the nature hypothesis. Van Gelder attempts to cope with this difficulty by means of a notion of *simulation* where he grants that a digital computer can simulate a dynamical system, by means of software which in

[17] For a similar point see Spurrett (1999:261f) concerning Cilliers' (1998) distinction between 'complex' systems, and those which are 'merely complicated'.

[18] This objection has been made to Van Gelder in the responses to his paper. See, e.g., Quartz (1998) and Beer (1998).

turn is implemented on the low level dynamical system which the computer instantiates (Van Gelder 1998:620). This long route from instantiation to implementation to simulation stands in stark contrast to the case of digital computers running abstract cognitive models which are themselves digital, where Van Gelder holds that the relationship between the computer and the model is one of *realisation*, given the stronger structural similarities between the model and the basic properties of the computer. Whether this answers the objection, and to what extent Van Gelder's proliferation of types of system-system relation (identity, instantiation, implementation, realisation, simulation) bears up under rigorous scrutiny are crucial questions, which we postpone until later in this paper. For now we need to note that Kirsh and Maglio's work definitely counts as a kind of exercise in what Van Gelder sees as the proper method of cognitive science, an instance of the 'knowledge hypothesis' at work and delivering results.

Kirsh and Maglio, though, do not deliver (for reasons noted above) a detailed working dynamical description of human Tetris playing. They describe a system which is both distributed and dynamical, to be sure, but it is what Van Gelder calls a description of 'how agents are causally organized at the highest level relevant to explanation of cognitive performances' (1998: 619). One might say that as the cognitive internalist model fell in the face of the facts, Kirsh and Maglio's ongoing research revealed the broad outlines of a distributed system with distributed boundaries and a particular division of labour across itself. This understanding permitted significant testable predictions.

Consider, for example, the action of rotating a zoid *before* it has completely emerged from the top of the screen, which Kirsh and Maglio observed to be common in human players. Recall the classical planning model of Tetris playing, and how on its terms early rotation, i.e. rotation before the computational problem has been 'solved' can have no possible value. Kirsh and Maglio suggest that these actions make sense however, if we consider the possibility of an epistemic function, namely that the rotations are enacted to unearth information early in the game. Given the advantage of early identification, and the fact that zoids only partially appearing from the top of the screen are on most occasions ambiguous, Kirsh and Maglio (1994) predicted that 'if a strategy exists for disambiguating shapes early then good players would strike on it'. Rotating a zoid early to expose initially hidden parts is such a strategy. Sometimes, however, rotations are not necessary to identify an emerging zoid, since if a player has expert knowledge of the game they will realise that certain shapes emerge only from certain columns and thus early rotation would reveal no new information. Interestingly, and in line with the thesis that early zoid rotation is epistemic, zoids that are completely ambiguous due to their initial partial images being identical in both shape and position, are rotated more than those which are not. Such a trend in the data is consistent with the thesis that the point of early rotation is to uncover new information immediately, since this bias in rotation rules out a purely pragmatic purpose. The action thus serves not to position the zoid towards the final pragmatic placement but rather to aid the computational end of revealing new information. Action and cognition are coupled.

The cognitive processing model which Kirsh and Maglio (1994) take their research to discover suggests that 'individual functional units inside the agent [are] in closed loop interaction with the outside world'. The result is a tight coupling between internal and external processes, which allows for 'offloading structure to the world, or for arranging things so that the world pre-empts the need for certain representations, or pre-empts the need for making certain inferences' (Kirsh & Maglio 1994). Action, on their view, is part of the *way in which cognition proceeds*.

It would be possible to explain in further detail the ways in which Kirsh and Maglio's work fits with the finer points of Van Gelder's account of dynamical systems, in particular his emphases on dynamicists' interest in changes rather than states, parallel rather than serial processes, ongoing processes rather than ones which divide neatly into input/output stages, caution or even hostility to the standard view of the role of internal representations in cognition, and the relative importance of 'coupling' compared to state-setting. Such a discussion would be partly tangential to our central purpose, though, and possibly of limited independent interest. The key point is that Kirsh and Maglio's work fits very well with Van Gelder's account of the knowledge hypothesis in cognitive science, and that it establishes a very plausible model of Tetris playing in which cognition is highly distributed, and where there are a variety of couplings between the agent as conceived from a cognitive internalist perspective and the body and environment of that agent. (As Van Gelder 1998:619 notes, the dynamical hypothesis is open to the possibility that a properly understood cognitive system 'might include variables not literally contained within the agent itself'.) This higher level dynamical understanding makes sense of the observational data, and is further supported by the range of subsidiary experiments which were part of Kirsh and Maglio's project.

There doesn't seem to be any problem with saying that the higher level dynamical functions and structures identified by Kirsh and Maglio are patterns in Dennett's sense. They are not observed from the physical stance, but relate to functions and regularities observable at the level of the cognitive system as a whole, and given both some input from the intentional and design stances. Are they *real* patterns, though?

Nature and Knowledge

Let us grant for the purposes of argument that Kirsh and Maglio satisfy the requirements of Van Gelder's knowledge hypothesis. The question, then, is whether there is a defensible inference in this case to the truth of the nature hypothesis—that Tetris players cannot just be understood as distributed dynamical systems, but that they *are* such systems. The couplings they describe between mind and world, brain and game could be seen as useful constructs in conceptualizing and accounting for the actions of Tetris players.

Van Gelder (1998:615) would presumably say that the answer here is 'yes'. He goes so far as to suggest that the dynamical hypothesis is a strong candidate for being the 'law of qualitative structure' for cognition, to replace Newell and Simon's proposal that cognition should be understood in terms of physical symbol systems. With reference to his differentiation of the dynamical hypothesis into the nature and knowledge hypotheses, Van Gelder claims that 'the best evidence for the former would be the truth of the latter' (Van Gelder 1998:619).

Just how good, though, is the best evidence here? Van Gelder is careful to note that even the *nature* hypothesis is 'concerned in the first instance not with low-level systems but with how agents are causally organised at the highest level relevant to an explanation of cognitive performance, whatever that may be' (1998: 619). So an understanding of a putative dynamical system in terms of the knowledge hypothesis would be a high-level dynamical model which, if correct, so that the nature hypothesis would be true in this case, would be *instantiated* by some real physical system. As noted above Van Gelder distinguishes instantiation from simulation, partly in order to make sense of how digital computers, whose qualitative structure is of the sort he wishes to say should not be taken as paradigmatic of cognition, can nonetheless be used to explore the dynamical systems he thinks are paradigmatic. Recall that the claim he made was that a digital computer can instantiate a simulation tool, which in turn can simulate a dynamical system of possible interest (Van Gelder 1998:620). Only the simulated system is considered to be a model of the target system on his view, rather than in what he calls orthodox computational modelling where *both* the concrete computational system and the abstract higher level digital system are considered models of the target system. The target system is in turn supposed be an instantiation of a dynamical system made out of concrete objects. (For good measure the target system in turn, for Van Gelder, realises an abstract dynamical system, to which it is hoped the system simulated on the concrete computer system will correspond. Happily this last layer is not crucial for our purposes.)

What should be manifest, if by no means obvious, is that the relationship between the target system and the concrete objects out of which it is composed is importantly similar to the relationship between an intentional system and the physical level of description on Dennett's account. The dynamical properties at the higher level of description are not of the same type as those at the basic level, they permit some variation between high and low level, which is to say that they can be multiply realised. This means that if we grant, for the purposes of argument, that the abstract dynamical system produced by some piece of cognitive science really does model a target system, even to the extent of being isomorphic with it, a question remains: Should we be realist or instrumentalist about the *target* system? If anything is to be considered an *abstracta*, it is, after all, the target system, a higher level dynamical system not identical with, and not reducible to, some set of concrete bottom level dynamics at the level of *Illiata*. So we're left on Dennett's knife-edge.

Ross (2000:147) quotes Dennett explaining why he is alarmed at philosophers who take 'Real Patterns' to be an exercise in metaphysics or ontology:

> I wouldn't want to trot out my ontology and then find I have to spend the rest of my life defending or revising it, instead of getting on with what to me are the genuinely puzzling issues—like the nature of consciousness, or selves, or free will (Dennett 1993:212).

For his part Ross thinks Dennett *does* need a metaphysics, and that 'Real Patterns' goes most of the way to delivering it, but ultimately fails to do so because of the way Dennett buys into the distinction between illiata and abstracta. He notes that the distinction can seem like a way of doing justice to the 'primacy of physics', which is clearly important for Dennett's materialism (Ross 2000:152).

Dennett is also, as we saw above, a deeply committed antireductionist. This cuts off one way of underwriting the illiata-abstracta distinction, since an ontological reductionist who is also convinced about the primacy of physics can view any non-fundamental kind as a posited kind (an abstracta) made up out of fundamental entities (the illiata). Apparent compromise positions, represented by a wide range of supervenience theories, founder on the fact that they are typically wedded to untenable assumptions about the nature of physics. As Ross (2000:155f) notes, supervenience theorists generally hold that science produces nomic generalisations, and that supervenient types can be multiply realised. Their antireductionism is based on the claim that the disjunctive set of physical tokens which is the base for a supervenient type will not be able to feature in the right kind of laws. But there is no reason to accept this view of how physics proceeds, and hence the required contrasts collapse, so that the supervenience theorist's 'defense of the reality of supervenient types threatens to collapse into instrumentalism: Our special sciences generalize over supervenient types *only* because we lack the epistemic resources necessary for identifying their disjunctive supervenience bases' (Ross 2000:156). Not only that, on this view the abstracta, or supervenient types, end up being anthropocentric—artefacts of *our* epistemic limitations.

Ross's argument (with which we are being very brief here) is that the way to proceed is to abandon the distinction between illiata and abstracta, and the implication of degrees of reality, or of first and second class ontological citizens, entirely. His proposed account of reality is as follows:

> To be is to be a real pattern, and a pattern is real if
> (i) it is projectable under at least one physically possible perspective, and
> (ii) it encodes information about at least one structure of events or entities S where that encoding is more efficient, in information theoretic terms, than the bit-map encoding of S, and where for at least one of the physically possible perspectives under which the pattern is projectable, there exists an aspect of S that cannot be tracked unless the encoding is recovered from the perspective in question (Ross 2000:161).

The chief strengths of this proposal are that it does justice to the primacy of physics, is not reductionist and also does not permit an infinite ontology. Crucially, it is not anthropocentric, and hence could not possibly be instrumentalist. We also note that it is partly an extension and refinement of the information theoretic notion of patterns hinted at in Dennett's own talk of compressibility in 'Real Patterns'. We don't aim to defend Ross's proposal here, at least not on its own metaphysical terms, but we do hope to show how by helping Dennett it helps Van Gelder and answers our question about the reality of the patterns, or 'structures', discovered by Kirsh and Maglio.

Van Gelder, like Dennett, needed a justification for the move from the applicability of some approach (dynamical systems, the intention stance) to realism about the type of structure to which that approach seemed to be committed. In both cases instrumentalism presented itself as either a genuine option, or an unacceptable alternative to be guarded against. Note that Ross's account gives neither Dennett nor Van Gelder the conclusion that the things which matter to them, intentional systems or cognitive-dynamical ones, actually *are* real. What it does do, though, is provide a way of saying that they *could* be, and what it would mean if they were. What exactly is or is not real is, quite properly, on Ross's view an empirical question, and one which there is no guarantee we would be able to answer. The answer to our question, then, is that what Kirsh and Maglio discover could indeed be a real pattern in a Dennettian sense. That is to say that research on distributed cognition could provide not just a powerful way (one among many) of *describing* cognition, but that it could bring us closer to a better understanding of the types of things there really are in the world.

Department of Philosophy
University of Natal

References

Bechtel, W & G Graham (eds) 1998. *A Companion to Cognitive Science.* Oxford: Blackwell.

Bechtel, W, A Abrahamsen & G Graham 1998. The Life of Cognitive Science. In W Bechtel, & G Graham (eds): *A Companion to Cognitive Science.* Oxford: Blackwell.

Beer, R 1998. Framing the Debate between Computational and Dynamical Approaches to Cognition. *Behavioral and Brain Sciences* 21:630.

Brooks, RA 1991a. Intelligence without Reason. *MIT AI Lab Memo* 1293, April 1991.

Brooks, RA 1991b. Intelligence without Representation. *Artificial Intelligence* 47:141-160.

Brooks, RA 1997. From Earwigs to Humans. *Robotics and Autonomous Systems* 20,2-4:291–304.

Cilliers, P 1998. *Complexity and Postmodernism: Understanding Complex Systems*. London: Routledge.

Clark, A 1989. *Microcognition: Philosophy, Cognitive Science and Paralled Distributed Processing.* Cambridge, Mass.: MIT Press.

Clark, A 1997. *Being There: Putting Brian, Body and World Together Again*. Cambridge, Mass.: MIT Press.

Dennett D 1978. *Brainstorms.* Cambridge, Mass.: MIT Press.

Dennett, D 1987. *The Intentional Stance.* Cambridge, Mass.: MIT Press

Dennett, D 1991. Real Patterns. *The Journal of Philosophy* 88:27-51.

Dennett, D 1993. Back from the Drawing Board. In Dahlbom, B (ed): *Dennett and His Critics*. Cambridge, Mass.: Blackwell.

Dennett, D 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. New York: Touchstone.

Dennett, D 1998. *Brainchildren: Essays on Designing Minds, 1984-1996.* Cambridge, Mass.: Bradford Books.

Dupré, J 1993. *The Disorder of Things*. Cambridge, Mass.: Harvard University Press.

Feyerabend, PK 1978. *Science in a Free Society*. London: Verso.

Feyerabend, PK 1987. *Farewell to Reason*. London: Verso.

Feyerabend, PK 1993. *Against Method (third edition)*. London: Verso.

Feyerabend, PK 1995. *Killing Time: The Autobiography of Paul Feyerabend*. Chicago: Chicago University Press.

Hutchins, E 1995. *Cognition in the Wild*. Cambridge, Mass.: MIT Press.

Huxley, A 1954. *The Doors of Perception*. London: Chatto & Windus.

Kim, J. 1998. *Mind in a Physical World*. Cambridge Mass.: MIT Press.

Kirsh, D & P Maglio 1991. Reaction and Reflection in Tetris. *Research Report D-015*, Cognitive Science Department, University of California, San Diego.

Kirsh, D & P Maglio 1992. Perceptive actions in Tetris, AAAI Spring Symposium on Control of Selective Perception, Stanford University.

Kirsh, D & P Maglio 1994. On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science* 18:513-549. Note pagination corresponds to the version at: http://cogsci.ucsd.edu/~kirsh/

Kirsh, D & P Maglio 1997. Some Epistemic Benefits of Action: Tetris, a Case Study. Proceedings of the fourteenth annual conference of the Cognitive Science Society, Hillsdale, New Jersey.

Koyré, A 1957. *From the Closed World to the Infinite Universe*. Baltimore: Johns Hopkins University Press.

Lorenz, EN 1963. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences* 20:130-141.

Maglio, PP, T Matlock, D Raphaely, B Chernicky, & D Kirsh 1999. Interactive Skill in Scrabble. In Proceedings of Twenty-first Annual Conference of the Cognitive Science Society. Mahwah, NJ: Lawrence Erlbaum.

Papineau, D 1993. *Philosophical Naturalism*. Oxford: Blackwell.

Quartz, S 1998. Distinguishing between the Computational and Dynamical Hypotheses: What Difference Makes the Difference? *Behavioral and Brain Sciences* 21:649-650.

Ross, D 2000. Rainforest Realism. In D Ross, A Brook, & D Thompson (eds): *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge, Mass.: MIT Press.

Ross, D, A Brook, and D Thompson (eds) 2000. *Dennett's Philosophy: A Comprehensive Assessment*. Cambridge, Mass.: MIT Press

Rumelhart, DE & JL McClelland (eds) 1986. *Parallel Distributed Processing. Volume I & II.* Cambridge, Mass.: MIT Press.

Spurrett, D 1999. Review Article of Cilliers, P,1998. Complexity and Postmodernism. London: Routledge. *South African Journal of Philosophy* 18,2:258-274.

Spurrett, D (this volume). The Case for Distributed Cognition: A Survey. *Alternation* 10,1.

Spurrett, D & Cowley, S (forthcoming). Putting Apes (Body and Language) Together again. To appear in *Language Sciences*.

Thelen, E, & L Smith 1994. *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, Mass.: MIT Press.

Valera, F, E Thompson and E Rosch 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, Mass.: MIT Press.

Van Gelder, T 1998. The Dynamical Hypothesis in Cognitive Science. *Behavioral and Brain Sciences* 21:615-665.

Vygotsky, LS 1968. Thought and Language. (Translation of 1962 edition). Cambridge, Mass.: MIT Press.

Wolfram, S 1983. Statistical Mechanics of Cellular Automata. *Rev. Mod. Phys*. 55:601-644.

Wolfram, S 1984. Universality and Complexity in Cellular Automata. *Physica D*. 10:1-35.