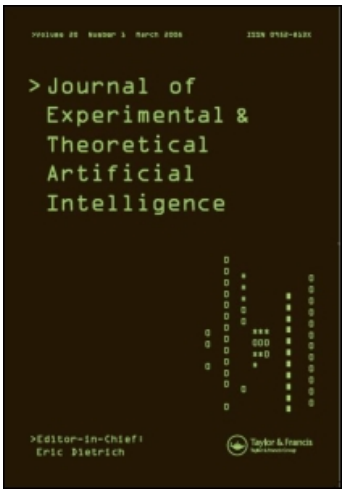

Journal of Experimental & Theoretical Artificial Intelligence
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713723652

On the nature of minds, or: truth and consequences

Shimon Edelman [a]
[a] Department of Psychology, Cornell University, Ithaca, NY, USA

Online Publication Date: 01 September 2008

To cite this Article Edelman, Shimon(2008)'On the nature of minds, or: truth and consequences',Journal of Experimental & Theoretical Artificial Intelligence,20:3,181 — 196
To link to this Article: DOI: 10.1080/09528130802319086
URL: http://dx.doi.org/10.1080/09528130802319086

Journal of Experimental & Theoretical Artificial Intelligence
Vol. 20, No. 3, September 2008, 181–196

TARGET ARTICLE

On the nature of minds, or: truth and consequences

Shimon Edelman*

Department of Psychology, Cornell University, Ithaca, NY, USA

(*Received 26 March 2008*)

Are minds really dynamical or are they really symbolic? Because minds are bundles of computations, and because computation is always a matter of interpretation of one system by another, minds are necessarily symbolic. Because minds, along with everything else in the universe, are physical, and insofar as the laws of physics are dynamical, minds are necessarily dynamical systems. Thus, the short answer to the opening question is 'yes'. It makes sense to ask further whether some of the computations that constitute a human mind are constrained by functional, algorithmic, or implementational factors to be essentially of the *discrete* symbolic variety (even if they supervene on an apparently continuous dynamical substrate). I suggest that here too the answer is 'yes' and discuss the need for such discrete, symbolic cognitive computations in communication-related tasks.

Keywords: computation; dynamical systems; symbolic dynamics; communication

1. On the nature of cognition

Minds are fundamentally computational phenomena (Turing 1950; McCulloch 1965; Marr 1982; Minsky 1985; Dennett 1991; McDermott 2001; Metzinger 2003; Minsky 2006; Nilsson 2006). The distinguishing and still surprisingly widely underappreciated feature of the explanatory framework for cognition that rests on this insight is that it has no viable alternatives (Edelman 2008).

Consider the so-called lightness problem, which is typical of visual perception: light from a source whose intensity is not directly known to the perceiver falls onto a surface whose reflectance needs to be estimated. The only course of action available to the perceiver is to try and estimate both the intensity and the reflectance from the amount of light that enters the measurement device (an eye or a camera). The pertinent observation here is that the essence of the process of distinguishing dark from light surfaces under potentially variable illumination — no matter how that process is instantiated — is a *computation*. Specifically, it consists of the factorisation of the number that represents the measured light quantity, whereby the source intensity and the surface reflectance that are multiplicatively confounded in the measurement are recovered (Edelman 2008, p. 8).[1] All other perceptual tasks can be given similarly explicit computational formulations.

Just as perception is inherently computational, so is action. As an example, consider the nature of the problem that is solved by a cat that jumps onto a table

*Email: se37@cornell.edu

ISSN 0952–813X print/ISSN 1362–3079 online

DOI: 10.1080/09528130802319086
http://www.informaworld.com

(Edelman 2008, p. 428). This problem consists of estimating the right amount of momentum that the cat must impart to itself — a momentum that would allow it to clear the table's near edge, yet would not send it sailing over the table to hit the wall on the far side. The cat's safe landing on the table indicates that its brain has successfully carried out a series of arithmetic operations on quantities representing (not necessarily explicitly in any sense, cf. Marr 1982, chap. 1) the acceleration of gravity on the cat's home planet, the cat's mass, the dimensions of the cat and of the table, and their relative positions.

These two examples, along with many others in perception, memory, language, thinking, and consciousness, all point to the same conclusion: no matter how else cognition can be *described*, computation is what it actually is (Edelman 2008). As a philosopher would put it, cognition is *realised* in computation.

2. On the nature of computation

This essential truth about cognition has an important consequence: *multiple realisability*. Because computation itself can be realised in different substrates — in a silicon chip, in neural tissue, in a clockwork mechanism, in a chemical network — so can cognition. A particular cognitive act — say, the factorisation of light intensity measurements into source and reflectance contributions — is the same, no matter how it is implemented; change the computation, and it will no longer be the same act. But what does it mean for two instances of computation to be the same? And what exactly is computation, anyway?

2.1. *What computation is*

A computation is a process that establishes a mapping among some symbolic domains. The mapping may take the form of a function or, more generally, a relation. The symbols may or may not be numbers: concluding that $3+2$ equals 5, determining that the intersection of the sets $\{c, e, k\}$ and $\{a, d, b, e, j\}$ is not empty, bundling together the concepts `cat`, `raccoon`, and `deer`, and convergence to the limit cycle of a system described by the equation $\mathrm{d}^2x/\mathrm{d}t^2 - \mu(1-x^2)\mathrm{d}x/\mathrm{d}t + x = 0$ are all perfectly good examples of computation.

Because it involves symbols, this definition is very broad: a system instantiates a computation if its dynamics can be interpreted (by another process) as establishing the right kind of mapping. It is possible for quite different systems to be computing the same thing. For example, an electrical circuit consisting of an inductor and a capacitor (a system described by Equation (1a) below) computes the same function of time as a brick suspended from a spring (Equation (1b)), provided that the charge held by the capacitor is interpreted as representing the brick's position, or vice versa:

$$V = -L\,\mathrm{d}^2q/\mathrm{d}t^2 = q/C \tag{1a}$$

$$F = m\,\mathrm{d}^2x/\mathrm{d}t^2 = -kx. \tag{1b}$$

Here, V is voltage across the capacitor, L and C are the inductance and the capacitance values, q is the electrical charge; F is the force applied to the spring, m is the mass of the brick, k is the spring's force constant and x is the brick's displacement. The solution of the first equation is $q = q_0\cos(\omega t)$, where $\omega = \sqrt{1/LC}$. The second one yields the same

time-dependent expression: $x = x_0 \cos(\omega t)$, where $\omega = \sqrt{k/m}$. It does not matter, of course, whether the quantity whose dependence on time is computed by either of those two systems is denoted by q or by x.

Under this definition, a stone rolling down a hillside computes its position and velocity in exactly the same sense that my notebook computes the position and the velocity of the mouse cursor on the screen (they just happen to be instantiating different symbolic mappings). Indeed, the universe in its entirety also instantiates a computation, albeit one that goes to waste for the lack of any process external to it that would make sense of what it is up to.

2.2. *When two computations are the same*

The notion that any physical system computes whatever one can plausibly claim it computes is a foundational one in computer science (Crutchfield 1994; McDermott 2001), as well as in computational neuroscience (Churchland and Sejnowski 1992).[2] Some critics (such as philosopher John Searle) question the explanatory value of computation for cognition in the light of its ubiquity. Their concern is based on the observation that given a snapshot of the state of a particular sufficiently complex system, a detailed correspondence can be drawn between its structure and the structure of some state of just about any conceivable computation. This implies, for example, that the instantaneous state of my brain at this moment may be isomorphic to that of some subset of the molecules of oxygen in the room.

The worry that a given complex system can be interpreted as implementing any arbitrarily chosen computation merely by following its natural dynamics is, however, unfounded. The combinatorial complexity of the universe ensures that step by step dynamic representation of one system by another that is plausible (i.e. both consistent and faithful) is a rarity. In particular, in the above example, the probability of my brain state and the state of the molecule ensemble just singled out remaining in strict correspondence as the dynamics of these two systems unfold over time is vanishing (Chalmers 1994; Block 1995). Thus, our foundational construal of computation does not entail complete and unfettered freedom of attributing computations to systems, because for such attribution to persist dynamically over time, the system in question and its computational description must undergo the same series of transitions between the putatively corresponding states.[3]

One process that is demonstrably capable of giving rise to such dynamical mimicry is evolution. Living systems are constantly under pressure to represent aspects of the world that are relevant to their survival (Dennett 1995; Clayton and Kauffman 2006). Some such systems do it by evolving natural computational devices that specialise in tracking the state of affairs of their dynamic environment, with the ultimate goal of anticipating it (Dennett 2003). These computational devices are brains.[4]

3. On the nature of explanation: levels of analysis

To fully explain an information processing device such as a brain, one must understand it on several levels of analysis (Marr and Poggio 1977; Marr 1982). At the most abstract, functional level, the challenge is to understand the problems for which computational solutions are needed. In the two examples mentioned in Section 1, these are, respectively, the factorisation of a product of two numbers into its constituents, and the solution of

a ballistic motion equation for a body in a constant gravitational field. At the next level, the concern is with the step-by-step procedures, no matter whether continuous or discrete, whereby a solution for the problem in question can be approached. Finally, at the bottom-most level one finds the possible implementations for the various computational procedures capable of solving the problem.[5]

The explanatory hierarchy broadens as one descends through functional, procedural, and implementation levels (Edelman 2008; chap. 4). For an animal, a certain behavioural need can usually be met in several functional ways, each of which leads to a distinct formulation of the problem that needs to be solved. Likewise, for each problem formulation, it is often possible to come up with several solution procedures. Finally, because a given computation can be realised in multiple distinct ways, a particular solution can be implemented in any of a number of substrates — e.g. in a neural computer, or an electronic one.

Although the proponents of hierarchical analysis of computational systems (notably, Marr and Poggio 1977) argued for it on epistemological grounds, it is also favoured by ontological considerations of the widest possible scope. The reason for this is that hierarchical structure is characteristic not just of systems that have evolved or have been engineered to carry out information processing, but rather of the universe in general:

> Scientific knowledge is organised in levels, not because reduction in principle is impossible, but because nature is organised in levels, and the pattern at each level is most clearly discerned by abstracting from the detail of the levels far below. [...] And nature is organised in levels because hierarchic structures – systems of Chinese boxes – provide the most viable form for any system of even moderate complexity.
>
> — Herbert A. Simon (1973, p. 26)

Simon's insight regarding the hierarchical structure of the universe is a cornerstone for developing a comprehensive understanding of embodied, situated brains that evolve to deal with the world (Edelman 2003b, 2008). When combined with the levels of analysis framework, this insight can be made to yield a comprehensive explanatory strategy for studying natural computation (and, therefore, cognition). Such a strategy has been outlined by Shalizi (2005):

> Many systems in statistical physics admit multiple levels of description, from microscopic molecular detail up through very broad macroscopic features. The higher-level descriptions are 'coarse-grainings' of the lower levels, and the higher-level variables are generally collective properties of many lower-level objects. Not every coarse-graining leads to a 'good' set of macroscopic variables; those that do have certain statistical properties. These properties, in turn, have important information-theoretic implications, and, when the coarse-graining is discrete ('symbolic dynamics'), the system can be modelled by stochastic automata.

To see how this approach 'might help cognitive scientists relate symbolic [...] descriptions to neural, dynamical ones' (Shalizi 2005; cf. Crutchfield 1994; Branicky 1995; Jaeger 1999; Smolensky 1999; beim Graben 2004; Dale and Spivey 2005), let us have a quick look at the two extreme views of computation that the idea of symbolic dynamics sets out to integrate.

3.1. *Discrete and continuous computation*

At the one extreme, there is the discrete model of computation formalised by Turing (1936). Its definition and properties are well-known and need not be rehearsed here (for an introductory overview, see Barker-Plummer 2007). At the other extreme, there is the

continuous dynamical system: 'a mathematical object that unambiguously describes how the state of some system evolves over time' (Beer 2000). Formally, a dynamical system is defined by a function that maps the present value of each of the system's state variables into a corresponding future value.

As Beer (2000) notes, differential equations and Turing machines are both examples of dynamical systems. The tension between the 'symbolic' and 'dynamical' traditions in cognitive science (Dale and Spivey 2005) that the present article sets out to mitigate (or at least to put into a proper methodological perspective) is, therefore, seen to arise from a kind of explanatory turf war. The disputed territory — cognition — is a set of phenomena that, being necessarily computational, would seem to be equally amenable to explanation in terms of the most general model of computation (the dynamical one), or in terms of the most intuitive one (Turing).

This dispute would have been moot if dynamical computation were more powerful than Turing computation. Although this is indeed the case in theory (e.g. the so-called analog shift map has computational power beyond the Turing limit; Siegelmann 1995), limitations on the precision of representation of real numbers that are required by super-Turing models prevent them from being implementable in practice (Schonbein 2005).[6] The choice, therefore, is really a matter of deciding which way of interpreting the behaviour (the input–output relations and the state evolution) of a cognitive system is better — that is, which one is more useful to the system that does the interpreting, be it a part of the brain that processes the output of another part or, as it were, a cognitive scientist studying the brain. In the latter case, a better interpretation is one that offers a more complete explanation. This brings us back to the issue of levels of explanation.

3.2. *The perils of neglecting levels of explanation*

Any explanation that leaves out the functional level (and is not readily relatable to a wider framework that does include the functional level) is incomplete in a more profound sense than an explanation that misses some implementational details. It is interesting to note, therefore, that it is the functional level that tends to be left out of the accounts of cognition — presumably because a theory-less mechanistic description is more easily mistaken for a complete explanation than an implementation-neutral functional theory.

The failure to distinguish algorithm- or implementation-level descriptions from complete theories is endemic in cognitive science writing. Consider, for example, Dale and Spivey's (2005, p. 320) phrase '... a particular theoretical framework, such as a production system or a connectionist model'. Contrary to what this expression implies, production systems and connectionist models are not 'theoretical frameworks': the former is an algorithmic approach to Turing-like computation, and the latter is an implementational style.

Connectionist (Feldman and Ballard 1982) theorising did play a useful role in cognitive science insofar as it sketched a possible alternative to theories rooted in symbolic logic and committed to a modular implementation inspired by the classical von Neumann computer architecture (e.g. Fodor 1975, 1983). In terms of the levels of analysis framework, the contribution of connectionism was to highlight the multiple possible ways of implementing a given function, as well as the interdependence between levels, as when the characteristics of the available implementational substrate constrain the computations that it can

perform, and therefore also the tasks that can be addressed (for an example from the domain of vision, see Edelman 1999, chap. 1).

These useful contributions have, however, been overshadowed by attempts to present connectionism as a 'miracle mind model' (Niklasson and Sharkey 1994), destined to obviate the need for classical theories of cognition. Connectionism itself is being elevated by its proponents to the status of a theory (e.g. of language, as in 'connectionist psycholinguistics'; Christiansen and Chater 2001). In some cases, near-miraculous explanatory powers are attributed not even to connectionism in general, but rather to specific connection schemes, such as 're-entry' (as noted by G.M. Edelman in his 2003 article). And yet, for a cognitive scientist to profess connectionism is the same as for an aeronautical engineer who would rather work on airplanes (as opposed to, say, helicopters) to avow 'fixedwingism'.

The explanatory value of these two stances in their respective fields is equally limited by virtue of their being confined to a single level of analysis. In the aeronautics example, a crucial piece of explanation — the Zhukovsky profile of the wing (fixed or rotary) that generates lift — is to be found on the level of airflow dynamics, not wing motility as such. Likewise, in brain science the explanatory keystone is typically found on the functional level, for instance, when deep parallels are revealed between popular 'connectionist' architectures and varieties of statistical computation (Omohundro 1987; Sarle 1994). In comparison, merely tracing the brain circuit that demonstrably implements some function explains very little about it, even if neuroanatomy is accompanied by a detailed compartmental model of the relevant neurophysiology.

It is easy to find similarly narrow views of explanation on the other side of the 'connectionism' *versus* 'classicism' debate. For instance, the influential critique of connectionism by Fodor and Pylyshyn (1988) is based on a category mistake that treats systematicity and compositionality as effectively synonymous (Edelman and Intrator 2003). Systematicity is a presumed desideratum in concept representation: it is claimed that to be useful, a representational scheme should be systematic in the sense that an agent capable of harbouring a relational representation $R(a, b)$ should also be capable of entertaining the notion that $R(b, a)$, independently of whether either or both of these relations happen to be true (Fodor and Pylyshyn 1988; Hadley 1994). One way of attaining systematicity is to make sure that representations are compositional: built up from smaller chunks, so that the meaning of the whole can be expressed as a function of the meaning of the parts. Now, the mistaken conflation of systematicity with compositionality hinges on a confusion between levels of explanation: whereas compositionality belongs to the procedural level, systematicity is a functional need (and an arguable one at that; see Johnson 2004).

A similar confusion underlies the busy ACT research programme (Anderson 1996), which boils down to the notion that the mind is a production system. As noted above, production systems are just one way of doing symbolic computation; an explanation that focusses on productions need not be wrong, but it cannot be complete. ACT's overcommitment to productions, and to the algorithmic level of explanation, which is where they belong, is illustrated by the nonchalant use that its 'connectionist' implementation (Lebiere and Anderson 1993) makes of embarrassingly implausible building blocks, such as 'neural networks' that operate on integer numbers in positional notation.

Some theories of cognition manage to leave out not one or two but all the levels of explanation that I have discussed so far, while attempting to link empirical behavioural findings directly to phenomenological narrative. A classical example of this approach is

radical 'direct perception' of Gibson (1979), who presaged the importance of embodiment and situatedness in cognition (an intellectual thread that is proving crucial in understanding cognition, and in particular development; Barsalou 1999; Smith and Gasser 2005), yet completely missed its computational essence (Ullman 1980).

I conclude this brief and necessarily cursory survey of overly narrow approaches to explanation by mentioning the research programme that took over its field by a storm and dominated it for half a century: formalist linguistics (see Chomsky (1995) for the current received version of the formalist theory of language, and Edelman and Waterfall (2007) for a review of its empirical status). Although some of its adherents refer to it as 'cognitive physiology' (Anderson and Lightfoot 2001) or 'biolinguistics' (Jenkins 2004) while stressing its computational origins (Chomsky 1957), formalist linguistics is equally dismissive of functional, algorithmic, and implementational (neural) 'details'; as Chomsky (2004, p. 56) recently put it, 'I think a linguist can do a perfectly good work in generative grammar without ever caring about questions of physical realism or what his work has to do with the structure of the mind'.

3.3. *An interim summary*

Given how evenly matched all the sides in the explanatory turf war are in their likelihood of shooting themselves in the foot, it is surprising how widespread the yearning still is out there for a kind of *Pax Romana* — whether dynamical, under which there would be no room for symbolic explanations, or symbolic, which would have it the other way around. The preceding discussion should make one equally leery of the older exclusively symbolic approach in the style of the 'language of thought' (Fodor 1975) and of the more recent 'dynamical systems' one (Port and van Gelder 1995) that aims to displace it, insofar as they strive for explanatory exclusivity while ignoring the multiple levels at which explanation is needed (for some examples, which these days are found mostly on the dynamical systems side of things, see Carello, Turvey, Kugler, and Shaw 1984; Gregson 1988; Thelen and Smith 1994; Van Orden 2002; Van Orden, Holden, and Turvey 2003; Turvey 2005).

Such ambitions are misguided for two reasons. First, as we just saw, full understanding requires consideration on several levels, among them functional, procedural and implementational. Second, as noted in Section 2, computation is in any case a matter of interpretation of one system by another (rather than an intrinsic property of any one system taken in isolation), and so are whatever symbols it may or may not involve. From this methodological vantage point, we can now have a fresh, and more informed, look at the question raised earlier in this section: whether or not one of the two competing models of computation, discrete and continuous, is better than the other in some circumstances.

4. Which kind of computation is more explanatory?

While addressing this question, we must keep in mind that for a computation to be really explanatory with respect to some cognitive function, rather than being merely a cognitive scientist's fancy, the system in question must of course be interpreted by another system (cf. Crutchfield 1994). The situation in which one system looks at the behaviour of another occurs naturally in the context of communication. There, as we shall see in this section, the discrete model of computation — more precisely, symbolic dynamics (Jaeger 1999;

Dale and Spivey 2005; Shalizi 2005; Spivey 2006) — enjoys a clear explanatory advantage (the similar need for discrete computation in other cognitive domains is defended by Dietrich and Markman (2003)).

The starting point for the following brief analysis is the observation that the behaviour of a given system may be assigned distinct (and even mutually incompatible) interpretations by interpreters with different computational abilities (Crutchfield 1994). Moreover, basing the process of discrete symbolic interpretation of a continuous dynamical system on a 'non-generating' partition of its state space can lead to self-inconsistent (let alone mutually incompatible) interpretations (beim Graben 2004).[7] These properties have methodological implications: they place epistemic limits on the application of symbolic dynamics in cognitive science. More importantly, however, these properties of computation in dynamical systems also have substantive implications: they give rise to constraints on the dynamics of computational elements whose behaviour needs to be interpreted by other components of the same system.

Reliable communication between two subsystems requires that the externally observable states of one of them be amenable to consistent and reliable interpretation by the other. Although it is not impossible to interpret continuous dynamical behaviour, systems whose components exhibit discrete symbolic dynamics enjoy several related advantages: (i) lower computational complexity, (ii) less sensitivity to noise, and (iii) better learnability.

4.1. *Managing computational complexity*

The possibility of transcribing speech in the form of sequences of distinct symbols that can subsequently be turned back into speech with very few errors demonstrates that natural language is discrete at a fundamental, pragmatically relevant level.[8] Discreteness, in turn, makes possible duality of patterning (Hockett 1960): meaning in language is conveyed to a large extent by combining elements that may themselves be meaningless. This familiar characteristic of natural speech is not a matter of necessity: in principle, communication can proceed through the exchange of continuous, parallel, globally modulated signals, with information being conveyed via the real-valued modulation parameters. In practice, however, natural language is discrete, serial, and locally informative.[9]

This choice of structure confers two related computational advantages: efficiency and open-ended expressiveness. The possibilities for encoding novel meanings in a discrete combinatorial language are open-ended because its semantic 'resolution' can always be increased in a trivial manner, simply by resorting to longer strings of symbols (in comparison, in continuous modulation it is limited by the resolution of the analog processor). Moreover, growth of message length can be kept in check by structuring the messages hierarchically, in a recursive manner.

All this suggests that there is an advantage to a discrete medium of communication, but what about the use made of this medium, namely, the coordination of internal representations employed by the systems that engage in communication? Communication between two agents is possible insofar as a morphism exists between their internal states. However, unless the two systems are clones, there is no single principled way of identifying states of one with states of the other via an isomorphism mapping.

This is just as well, because the meaning of an internal state (which may or may not be linked to an external state of affairs) for the system itself is most naturally defined in terms

of that state's relations to its other states (Edelman 1998). This implies that communication between two differently structured systems is still possible if partial *second-order isomorphism* (Shepard and Chipman 1970) is established between the representations they harbour individually (Edelman 1999); that is, if relations between states of one system are mapped to relations between states of the other (Edelman 1999; Goldstone and Rogosky 2002).

Second-order isomorphism, in turn, requires a principled correspondence between various states of a given system. In a continuous state space, this means something like mapping trajectories through R^n to themselves. The complexity of this problem, as captured by the number of possible matches for each point, is intractable, unless constraints are imposed to regularise it. The most straightforward way to do so is by discretising the state space, by constraining the system to operate in a symbolic dynamics regime (Jaeger 1999; Dale and Spivey 2005; see also, Section 5).

But do we really need symbolic dynamics here? It would seem that any two continuous dynamical systems that are coupled together (weakly, lest they become effectively one) can be said to communicate with one another. This account of communication, if accepted at the face value, illustrates all that is wrong with the radical version of dynamical systems theorising: it happens to be valid on the implementation level, and yet it is woefully inadequate, because it leaves the key computational-level issue — the nature of the principled correspondence between the states of the two systems — unresolved.

4.2. *Resisting noise*

For a continuous state-space representation to be robust against noise means to have 'safety margins' around each trajectory in those places where it passes close to other trajectories, with which it must not be confused. When desensitised to noise in this manner, a continuous state space is effectively discretised (indeed, as noted in the previous section, this is precisely how the 'super-Turing' computational model that relies on real-number representations becomes Turing-equivalent when implementational constraints are factored in; Schonbein 2005).

Having been reduced to implementing symbolic dynamics (with each safety region playing the role of a 'symbol'), a dynamical system becomes capable not just of error tolerance but of error correction through redundant coding. Understandably, the most obvious instance of biological information processing systems resorting to discrete representations is the bottleneck through which their specifications must pass to reach the next generation: the genetic code.

4.3. *Learning*

In species capable of learning through communication, language (or song) is another medium through which information can bridge generations. In addition to making communication feasible, representations that are effectively discrete also make it much easier to learn to communicate. In an analog communication system that uses a global orthogonal basis to span the space of possible signals, the learnability of the basis poses a serious computational issue: given a sample of signals, how to choose the right basis among the uncountable infinity of possible ones.

This problem does not arise in a system that uses a discrete, local, serially compositional basis for language: the basic symbols can be learned by the discovery method outlined by Harris (1946), which, when applied recursively with the appropriate statistical checks and balances on structure inference, leads to the possibility of bootstrapping syntax from experience (Solan, Horn, Ruppin, and Edelman 2005; Sandbank, Berant, Edelman, and Ruppin 2008). Evolutionary simulations suggest that the advantages conferred by discrete compositionality are such that 'compositional syntax is an inevitable outcome of the dynamics of observationally learned communication systems' (Kirby 2000, p. 303).

This conclusion is supported by developmental data, and by observations of languages that emerge and evolve even as they are being studied by linguists. A longitudinal study of the Nicaraguan Sign Language, which focussed among other characteristics on discreteness and combinatorial patterning, found that within a few decades the representation of motion shifted from overlapped and analog to sequential and digital (Senghas, Kita, and Özyürek 2004). Similarly, earlier observations of the acquisition of American Sign Language (ASL) showed that children learning ASL develop a preference for linear sequencing even in cases where adults use overlapped constructions (Newport 1981).

5. How a continuous state space gets its spots

In the preceding sections, we saw that for a dynamical system to support the emergence of tractable and stable mappings between states, and for its state transitions to be amenable to interpretation by other systems, it must, at a certain level of description, exhibit 'symbolic' (quasi-discrete) dynamics. At that level of description, some of the states in the system in question serve as *dynamical symbols*: 'intrinsically isolatable, intrinsically classifiable events' (Jaeger 1999), whose beginning and end must stand out from the background fluctuations, so that it is possible to tell two symbols apart without relying on fine tuning of measurement parameters.

This latter property reduces the degree of arbitrariness of the computation attributed to the system.[10] The discrete computation embodied in the symbolic dynamics will be generic (cf. Freeman 1993), that is, the same for an entire class of interpretation procedures. This in turn will make it a quantifiably good explanation of the system's behaviour, in a sense that is related to Occam's Razor and to Bayesian model inference (MacKay 1991, 2003).

The dynamical phenomenon that fits the need for intrinsic identifiability is transient attractor — a region of the state space defined by the volume contraction property. Think of a bundle of trajectories through the state space that are close to each other (semi-formally, this means that they have just squeezed together through a compact Poincaré section). Upon approaching a transient attractor, such a bundle pulls even tighter, leading to a contraction of the state space volume that it delineates (Jaeger 1999, Figure 2; cf. Spivey 2006, Figures 12.1, 12.2; note that a permanent rather than transient attractor would not do here, because the system will never leave after falling into one).

Jaeger (1999) doubts that defining symbolic dynamics in terms of transient attractors is general enough in the light of the great potential diversity of the dynamics of brains embedded in a complex ecosystem.[11] It seems to me, however, that natural selection, rather than invariably disrupting symbolic dynamics, is just as likely to sustain and propagate it.

Agents that stand to benefit from communication, or even just from the kind of mutual social simulation that underlies theory of mind reasoning (Gallese and Goldman 1998; Grush 2004), are subject to evolutionary pressure towards harbouring mutually interpretable representations.

It is worth noting that a system that needs to communicate something *to itself* at a later time can benefit from precisely the same functionality as offered by transient attractors. One way for a brain to send a message to itself in the future is, of course, to lay down a long-term memory trace — a function subserved in primates by the hippocampus and the posterior cortical areas (Merker 2004). But what if the storage must be temporary, as in problem solving, which requires working memory (a 'scratchpad'), or more generally in complex action planning, which requires that a representation of the goal be maintained in the face of changing context?

A transient attractor that is temporarily isolated from the rest of the system's dynamics fits the bill. Indeed, primate behaviours related to working memory, planning, and perhaps also language, rely on 'digital' representations maintained in the prefrontal cortex and controlled by reward-dependent gating signals ascending from the basal ganglia (O'Reilly 2006).

6. Conclusions

Given the explanatory reduction[12] of minds to computations (outlined in Section 1; see Edelman (2008) for a detailed and comprehensive treatment), gaining a clear understanding of the nature of computation is crucial for understanding the mind. This truth has certain consequences for cognitive science.

6.1. *The truth*

A system of which computation is the essence can only be understood if considered simultaneously on several levels of analysis. On the abstract levels of problem and of procedure, the 'discrete symbols *versus* continuous dynamics' debate is strictly irrelevant. These levels deal with computation as abstracted away from any particular implementation — a step that is always possible (even if the actual procedural solution is dictated in part by implementational constraints), because computation is, as a matter of principle, multiply realisable. Thus, insofar as various models of computation (discrete or continuous) have the same power (a notion that can be made formal), they are all equivalent, and the 'dynamical systems' issue is moot.

On the fundamental physical implementation level, one question that arises is which of the multiple possible abstractions offers the best explanation for the functioning of a given cognitive system. This issue is epistemological and as such seems to pertain to the methodology of cognitive science rather than to its subject matter. From the epistemological standpoint, the conceptual toolbox borrowed from continuous dynamical systems analysis would be trivially applicable to any system[13] — except that in practice it gets bogged down in intractability stemming from the high dimensionality of the state space in interesting systems such as brains. Thus, in practice, looking at raw brain dynamics seems only to yield mass-action 'laws' (e.g. Van Orden, Holden, and Turvey 2003) that are descriptive rather than explanatory. Moreover, such descriptions apparently

miss the most important aspects of brain function (Bullock 2003). In comparison, the conceptual toolbox of discrete computation is much more tolerant of complexity.[14]

The upshot of Section 4 that I still find somewhat surprising is that the seemingly purely methodological issue of how to interpret the behaviour of a cognitive system happens also to have substantive implications: there is a matter of fact about what cognitive systems actually do, independently of what the scientists who study them might believe they do. Specifically, the epistemological considerations apply intrinsically to a collection of interacting systems, each of which is charged with interpreting the behaviour of others (in the same sense that a scientist may attempt to understand their workings). This fact opens the door to the possibility of some but not other modes of computation being *intrinsically* advantageous for certain systems or tasks. The rough sketch of an analysis offered in this article suggests that this is indeed the case: discrete computation is more appropriate for communication-like situations.

6.2. *The consequences*

One may argue for explanatory diversity in cognitive science by observing that it is a good idea because cognition is computation, because computation is always a matter of interpretation, because interpretation depends on the level of analysis, because discrete and continuous accounts of computation often pertain to different levels, and because explanations of complex phenomena tend to span more than one such level. This rather familiar line of reasoning is now supplemented by a new argument that builds a bridge from considerations of scientific epistemology to intrinsically epistemic (and in that sense, ontic) constraints on certain cognitive systems. For such systems, being digital is a real need. In a setting whose dynamics is, deep down inside, effectively continuous (as it seems to be the case with neurons at the level of membrane voltages and ion currents), the way to fulfill this need is to resort to symbolic dynamics.

Acknowledgement

Many thanks to Rick Dale and to Barb Finlay for detailed and insightful comments on a draft of this article.

Notes

1. How two unknowns can be recovered from a single measurement is an interesting question (Marr 1982; Edelman 2008) that is beside the point for the present discussion.
2. And perhaps even in physics and cosmology (Wolfram 2002).
3. The isomorphism between state transitions does not have to unfold in real time; cf. Grush 2004.
4. A mind is not confined to an individual's brain; rather, it spills over into the environment and in particular, into other individuals' brains (Dennett 2003, p. 122).
5. For an early proposal for a multi-level explanation of how the vertebrate retina may be solving the lightness problem, see Marr (1974). In this connection, see also the multi-level discussion of why the chicken crossed the road (Edelman 2008, table 4.1).
6. Note that quantum computation, which is super-Turing, is most probably not relevant to cognition (Tegmark 2000; Koch and Hepp 2006), even if it proves to be feasible.
7. A partition is 'generating' if refining it by resorting to progressively longer discrete representations allows arbitrarily small intervals of initial conditions to be singled out (Crutchfield 1994).

8. It may be continuous at other levels, as it was shown by Spivey (2006) for the case of processing of lexical items, for example; still, such continuous processing is periodically checked by discrete 'milestones' — just as a skier rushing down the slope ascribes a continuous trajectory, yet must pass either to the left or to the right of the tree that blocks her way.
9. To make this work, the segments that comprise the communication signal as it unfolds over time must, of course, be sufficiently distinct. This is true of spoken languages despite coarticulation (and the supposedly 'blurry' phonology; Port and Leary 2005), and even of sign languages (Sandler 2006).
10. It is worth reiterating here that given a particular process, what is up to interpretation is not whether or not it constitutes a computation, but rather *which* computation it constitutes. As noted in Section 2.2, multiple answers are usually possible, but their number decreases with the complexity of the process and of the computational interpretation that is being attributed to it.
11. In this connection, Jaeger (1999) writes, 'I [. . .] believe that neural dynamics, propped by billions of years of freewheeling evolution, intrinsically defies clean definitions — mathematical rigour hardly being a fitness criterion in natural selection. As a consequence, I do not think that the notion of dynamical symbols can be mathematically defined'.
12. For a definition of explanatory reduction, see Dennett (1995, p. 195).
13. Unless Lloyd (2008) is right and 'the universe [is], at bottom, digital'.
14. This is why the computer on which I am writing these lines is digital rather than analog. Even the kind of oscilloscope that I used as an undergraduate in the dynamical systems and control lab in the engineering school to plot limit cycles would these days be wholly digital, behind its analog-to-digital front end.

References

Anderson, J.R. (1996), 'ACT: A Simple Theory of Complex Cognition', *American Psychologist*, 51, 355–365.

Anderson, S.R., and Lightfoot, D.W. (2001), *The Language Organ: Linguistics as Cognitive Physiology*, Cambridge, UK: Cambridge University Press.

Barker-Plummer, D. (2007), 'Turing Machines', in *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta, *http://plato.stanford.edu/archives/win2007/entries/turing-machine/*

Barsalou, L.W. (1999), 'Perceptual Symbol Systems', *Behavioral and Brain Sciences*, 22, 577–660.

Beer, R.D. (2000), 'Dynamical Approaches to Cognitive Science', *Trends in Cognitive Sciences*, 4, 91–99.

beim Graben, P. (2004), 'Incompatible Implementations of Physical Symbol Systems', *Mind and Matter*, 2, 29–51.

Block, N. (1995), 'The Mind as the Software of the Brain', in *AnInvitation to Cognitive Science*, eds. D.N. Osherson, L. Gleitman, S.M. Kosslyn, S. Smith and S. Sternberg, Cambridge, MA: MIT Press.

Branicky, M.S. (1995), 'Universal Computation and other Capabilities of Hybrid and Continuous Dynamical Systems', *Theoretical Computer Science*, 138, 67–100.

Bullock, T.H. (2003), 'Have Brain Dynamics Evolved? Should we look for Unique Dynamics in the Sapient Species?', *Neural Computation*, 17, 2013–2027.

Carello, C., Turvey, M.T., Kugler, P.N., and Shaw, R.E. (1984), 'Inadequacies of the Computer Metaphor', in *Handbook of cognitive neuroscience*, ed. M.S. Gazzaniga, New York: Plenum, pp. 229–248.

Chalmers, D. (1994), 'On Implementing a Computation', *Minds and Machines*, 4, 391–402.

Chomsky, N. (1957), *Syntactic Structures*, Mouton: the Hague.

Chomsky, N. (1995), *The Minimalist Program*, Cambridge, MA: MIT Press.

Chomsky, N. (2004), *The Generative Enterprise Revisited*, Berlin: Mouton de Gruyter (Discussions with Riny Huybregts, Henk van Riemsdijk, Naoki Fukui and Mihoko Zushi).

Christiansen, M.H., and Chater, N. (2001), 'Connectionist Psycholinguistics: Capturing the Empirical Data', *Trends in Cognitive Sciences*, 5, 82–88.
Churchland, P.S., and Sejnowski, T.J. (1992), *The Computational Brain*, Cambridge, MA: MIT Press.
Clayton, P., and Kauffman, S.A. (2006), 'Agency, Emergence, and Organisation', *Biology and Philosophy*, 21, 501–521.
Crutchfield, J.P. (1994), 'The Calculi of Emergence: Computation, Dynamics, and Induction', *Physica D*, 75, 11–54.
Dale, R.A., and Spivey, M.J. (2005), 'From Apples and Oranges to Symbolic Dynamics: A Framework for Conciliating Notions of Cognitive Representation', *Journal of Experimental & Theoritical Artificial Intelligence*, 17, 317–342.
Dennett, D.C. (1991), *Consciousness Explained*, Boston, MA: Little, Brown & Company.
Dennett, D.C. (1995), *Darwin's Dangerous idea: Evolution and the Meanings of Life*, New York: Simon & Schuster.
Dennett, D.C. (2003), *Freedom Evolves*, New York: Viking.
Dietrich, E., and Markman, A.B. (2003), 'Discrete Thoughts: Why Cognition must use Discrete Representations', *Mind and Language*, 18, 95–119.
Edelman, G.M. (2003a), 'Naturalizing Consciousness: A Theoretical Framework', *Proceedings of the National Academy of Science*, 100, 5520–5524.
Edelman, S. (1998), 'Representation is Representation of Similarity', *Behavioral and Brain Sciences*, 21, 449–498.
Edelman, S. (1999), *Representation and Recognition in Vision*, Cambridge, MA: MIT Press.
Edelman, S. (2003b), 'But will it Scale up? not without Representations', *Adaptive Behavior*, 11, 273–275 (A Commentary on the Dynamics of Active Categorical Perception in an Evolved Model Agent by R. Beer).
Edelman, S. (2008), *Computing the Mind: How the Mind Really Works*, New York: Oxford University Press.
Edelman, S., and Intrator, N. (2003), 'Towards Structural Systematicity in Distributed, Statically Bound Visual Representations', *Cognitive Science*, 27, 73–109.
Edelman, S., and Waterfall, H.R. (2007), 'Behavioral and Computational Aspects of Language and its Acquisition', *Physics of Life Reviews*, 4, 253–277.
Feldman, J.A., and Ballard, D.H. (1982), 'Connectionist Models and their Properties', *Cognitive Science*, 6, 205–254.
Fodor, J. (1975), *The Language of Thought*, New York: Crowell.
Fodor, J., and Pylyshyn, Z. (1988), 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, 28, 3–71.
Fodor, J.A. (1983), *The Modularity of Mind*, Cambridge, MA: MIT Press.
Freeman, W.T., (1993), 'Exploiting the Generic view Assumption to Estimate Scene Parameters', in *Proceedings of the 3rd International Conference on Computer Vision*, Washington: IEEE, pp. 347–356.
Gallese, V., and Goldman, A. (1998), 'Mirror Neurons and the Simulation Theory of Mind-Reading', *Trends in Cognitive Sciences*, 12, 493–501.
Gibson, J.J. (1979), *The Ecological Approach to Visual Perception*, Boston, MA: Houghton Mifflin.
Goldstone, R.L., and Rogosky, B.J. (2002), 'Using Relations Within Conceptual Systems to Translate Across Conceptual Systems', *Cognition*, 84, 295–320.
Gregson, R.A.M. (1988), *Nonlinear Psychophysical Dynamics*, Hillsdale, NJ: Erlbaum.
Grush, R. (2004), 'The Emulation Theory of Representation: Motor Control, Imagery, and Perception', *Behavioral and Brain Sciences*, 27, 377–442.
Hadley, R.F. (1994), 'Systematicity Revisited', *Mind and Language*, 9, 431–444.
Harris, Z.S. (1946), 'From Morpheme to Utterance', *Language*, 22, 161–183.
Hockett, C.F. (1960), 'The Origin of Speech', *Scientific American*, 203, 88–96.

Jaeger, H. (1999), 'From Continuous Dynamics to Symbols', in *Dynamics, Synergetics, Autonomous Agents*, eds. W. Tschacher and J.-P. Dauwalder, Singapore: World Scientific, pp. 29–48.
Jenkins, L. (Ed.) (2004), *Variation and Universals in Biolinguistics*, Vol. 62 of North-Holland Linguistic Series: *Linguistic Variations*, Elsevier: Amsterdam.
Johnson, K.E. (2004), 'On the Systematicity of Language and Thought', *Journal of Philosophy*, CI:3, 111–139.
Kirby, S. (2000), 'Syntax without Natural Selection: How Compositionality Emerges from Vocabulary in a Population of Learners', in *The Evolutionary Emergence of Language*, eds. C. Knight, M. Studdert-Kennedy and J.R. Hurford, Cambridge: Cambridge University Press, pp. 303–323.
Koch, C., and Hepp, K. (2006), 'Quantum Mechanics in the Brain', *Nature*, 440, 611–612.
Lebiere, C., and Anderson, J.R. (1993), 'A Connectionist Implementation of the ACT-r Production System', in *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pp. 635–640.
Lloyd, S. (2008), 'Quantum Information Matters', *Science*, 319, 1209–1211.
MacKay, D.J.C. (1991), 'Bayesian Methods for Adaptive Models', Ph.D. thesis, California Institute of Technology, Pasadena, CA.
MacKay, D.J.C. (2003), *Information Theory, Inference, and Learning Algorithms*, Cambridge, UK: Cambridge University Press.
Marr, D. (1974), 'The Computation of Lightness by the Primate Retina', *Vision Research*, 14, 1377–1388.
Marr, D. (1982), *Vision*, San Francisco, CA: W. H. Freeman.
Marr, D., and Poggio, T. (1977), 'From Understanding Computation to Understanding Neural Circuitry', *Neurosciences Research Program Bulletin*, 15, 470–488.
McCulloch, W.S. (1965), *Embodiments of Mind*, Cambridge, MA: MIT Press.
McDermott, D.V. (2001), *Mind and Mechanism*, Cambridge, MA: MIT Press.
Merker, B. (2004), 'Cortex, Countercurrent Context, and Dimensional Integration of Lifetime Memory', *Cortex*, 40, 559–576.
Metzinger, T. (2003), *Being no one: The Self-Model Theory of Subjectivity*, Cambridge, MA: MIT Press.
Minsky, M. (1985), *The Society of Mind*, New York: Simon and Schuster.
Minsky, M. (2006), *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind*, New York: Simon & Schuster.
Newport, E.L. (1981), 'Constrains on Structure: Evidence from American Sign Language and Language Learning', in *Aspects of the Development of Competence,* Vol. 14 of Minnesota Symposia on Child Psychology, ed. W.A. Collins, Hillsdale, NJ: Erlbaum, pp. 93–124.
Niklasson, L.F., and Sharkey, N. (1994), 'Connectionism – the Miracle Mind Model', in *Connectionism in a Broad Perspective*, eds. L.F. Niklasson and M.B. Bodén, Chichester, UK: Ellis Horwood, pp. 13–25.
Nilsson, N.J. (2006), *How are we to know?*, *http://ai.stanford.edu/~1nilsson/hawtk/hawtk-webpage.htm*
Omohundro, S.M. (1987), 'Efficient Algorithms with Neural Network Behaviour', *Complex Systems*, 1, 273–347.
O'Reilly, R.C. (2006), 'Biologically based Computational Models of High-Level Cognition', *Science*, 314, 91–94.
Port, R.F., and Leary, A.P. (2005), 'Against Formal Phonology', *Language*, 81, 927–964.
Port, R.F., and van Gelder, T. (Eds.) (1995), *Mind as Motion: Explorations in the Dynamics of Cognition*, Cambridge, MA: MIT Press.
Sandbank, B., Berant, J., Edelman, S., and Ruppin, E. (2008), *From Context to Grammar: Inferring Rich Grammatical Structure from Raw Text* (Submitted).
Sandler, W. (2006), 'An Overview of Sign language Linguistics', in *Encyclopedia of Language and Linguistics* (Vol. 11, 2nd ed.) ed. K. Brown, Amsterdam: Elsevier, pp. 328–338.

Sarle, W.S., (1994), 'Neural Networks and Statistical Models', in *Proceedings of the 19th Annual SAS Users Group International Conference*, Cary, NC: SAS Institute, pp. 1538–1550.
Schonbein, W. (2005), 'Cognition and the Power of Continuous Dynamical Systems', *Minds and Machines*, 15, 57–71.
Senghas, A., Kita, S., and Özyürek, A. (2004), 'Children Creating Core Properties of Language: Evidence from an Emerging Sign Language in Nicaragua', *Science*, 305, 1779–1782.
Shalizi, C.R., (March, 2005), 'Symbolic Dynamics, Coarse-Graining, and Levels of Description in Statistical Physics and Cognitive Science', in *Proceedings of Workshop on Symbol Grounding: Dynamical Systems Approaches to Language*, Potsdam, pp. 14–17.
Shepard, R.N., and Chipman, S. (1970), 'Second-Order Isomorphism of Internal Representations: Shapes of States', *Cognitive Psychology*, 1, 1–17.
Siegelmann, H.T. (1995), 'Computation Beyond the Turing Limit', *Science*, 268, 545–548.
Simon, H.A. (1973), 'The Organisation of Complex Systems', in *Hierarchy Theory: The Challenge of Complex Systems*, Chap. 1, ed. H.H. Pattee, New York: George Braziller, pp. 1–28. Chapter 1.
Smith, L.B., and Gasser, M. (2005), 'The Development of Embodied Cognition: Six Lessons from Babies', *Artificial Life*, 11, 11–30.
Smolensky, P. (1999), 'Grammar-based Connectionist Approaches to Language', *Cognitive Science*, 23, 589–613.
Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2005), 'Unsupervised Learning of Natural Languages', *Proceedings of the National Academy of Science*, 102, 11629–11634.
Spivey, M.J. (2006), *The Continuity of Mind*, New York: Oxford University Press.
Tegmark, M. (2000), 'Importance of Quantum Decoherence in Brain Processes', *Physical Review E*, 61, 4194–4206.
Thelen, E., and Smith, L.B. (Eds.) (1994), *A Dynamic Systems Approach to the Development of Cognition and Action*, Cambridge, MA: MIT Press.
Turing, A.M. (1936), 'On Computable Numbers, with an Application to the Entscheidungs Problem', in *Proceedings of the London Mathematical Society, Series 2*, 42, 230–265.
Turing, A.M. (1950), 'Computing Machinery and Intelligence', *Mind*, 59, 433–460.
Turvey, M.T. (2005), 'Theory of Brain and Behaviour in the 21st Century: No Ghost, No Machine', *Japanese Journal of Ecological Psychology*, 2, 69–79.
Ullman, S. (1980), 'Against Direct Perception', *Behavioral and Brain Sciences*, 3, 373–416.
Van Orden, G., Holden, J., and Turvey, M.T. (2003), 'Self-Organisation of Cognitive Performance', *Journal of Experimental Psychology: General*, 132, 331–351.
Van Orden, G.C. (2002), 'Nonlinear Dynamics and Psycholinguistics', *Ecological Psychology*, 14, 1–4.
Wolfram, S. (2002), *A New Kind of Science*, Champaign, IL: Wolfram Media.