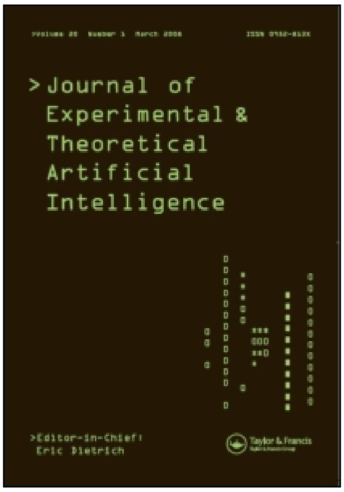

Journal of Experimental & Theoretical Artificial Intelligence
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713723652

On a compatibility between emergentism and reductionism

Michael J. Spivey [a]; Sarah E. Anderson [b]
[a] Department of Cognitive Science, University of California, Merced, USA; [b] Department of Psychology, Cornell University, Ithaca, USA

Online Publication Date: 01 September 2008

To cite this Article Spivey, Michael J. and Anderson, Sarah E.(2008)'On a compatibility between emergentism and reductionism',Journal of Experimental & Theoretical Artificial Intelligence,20:3,239 — 245

To link to this Article: DOI: 10.1080/09528130802319193

URL: http://dx.doi.org/10.1080/09528130802319193

Journal of Experimental & Theoretical Artificial Intelligence
Vol. 20, No. 3, September 2008, 239–245

COMMENTARY

On a compatibility between emergentism and reductionism

Michael J. Spivey[a*] and Sarah E. Anderson[b]

[a]*Department of Cognitive Science, University of California, Merced, USA;*
[b]*Department of Psychology, Cornell University, Ithaca, USA*

To the overarching question in this special issue as to whether it is wise to pursue an inclusive pluralist approach to dynamical and symbolic theoretical frameworks in cognitive science, Rick Dale answers 'Yes, because everybody's right,' Shimon Edelman answers, 'Yes, because everybody's wrong,' and David Jilk and colleagues declare, 'Hey, look what we built!' But seriously, the proof of the pudding is in the eating, as Don Quixote's squire Sancho once said, and Jilk, Lebiere, O'Reilly, and Anderson (2008) have put forth a very consumable existence proof in their hybrid SAL model. While they do devote some text to the philosophical issues surrounding scientific pluralism, they clearly prefer proof of concept over logical argumentation. Duct tape jokes aside, their combined production-system-cum-neural-network performs impressively on a range of tests and thus clearly warrants enthusiasm. That said, we nonetheless feel compelled to deliver in this commentary a cautionary note regarding the use of discrete symbols and formal rules in any model of cognition.

1. Conservative emergentism

In fact, we find ourselves in the awkward position of arguing for exactly that which Dale (2008) calls 'unrealistic,' Jilk et al. (2008) dismiss as 'reductionism,' and Edelman (2008) likens to 'fixed-wingism' (an arbitrary preference for an airplane's lift over a helicopter's lift).[1] Although we are certainly not opposed to cooperative interdisciplinary research in cognitive science, we feel strongly that for most (if not all) of perception and cognition, there is one particular unifying framework that is providing the greatest combination of success and promise: a complex dynamical systems approach (fuelled also by ecological psychology, embodied cognition and connectionism). We call our philosophical position here a 'conservative emergentism' – because neither 'liberal reductionism' nor 'unrealistic reductionist fixed-wingism' sounded good to us. Conservative emergentism has some overlap with what Emmeche, Køppe, and Stjernfelt (2000) refer to as 'weak downward causation'. It is a simultaneous attempt to: (a) acknowledge that global coherent structures emerge somewhat unpredictably in complex non-linear systems and (b) maintain the concerted effort to reveal the dynamic physical underpinnings of those emergent properties. Curiously, this intermediate position

*Corresponding author. Email: spivey@cornell.edu

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

DOI: 10.1080/09528130802319193
http://www.informaworld.com

puts us not so much in between the two warring sides (of symbolic and dynamical cognition), but instead in the 'peanut gallery', complaining as the two sides find out that they agree with one another far more than they had realised.

Classical cognitive science has traditionally relied on a unidirectional feedforward subsymbolic-to-symbolic account of mind (much like the sequence of events in a symbolic dynamics simulation) where practically all the emphasis is placed at the level of the psychological outcomes, rather than the underlying subsymbolic processes. On the other extreme of this war, radical emergentist cognitive science is relying on global coherent cognitive patterns as emergent properties that are not predictable from or decomposable to, their underlying physical elements – emergent properties which may in fact themselves alter the dynamics of their own underlying physical elements (but cf. Kim 2000).[2] Thus, they too are placing practically all of their emphasis at the level of the psychological outcomes, rather than the underlying subsymbolic processes.

It is empirically demonstrated that neural feedback from (high-level) cognitive categories can quickly influence the underlying dynamics of (low-level) perception (e.g. Ganis, Thompson, and Kosslyn 2004; Lupyan and Spivey 2008; Puri and Wojciulik 2008). Therefore, we are not in principle opposed to a weak form of the radical emergentist's downward causation (e.g. Emmeche et al. 2000). However, it may not be necessary to treat those cognitive categories as describable only at the higher level of organisation (cf. McRae, de Sa, and Seidenberg 1997; Rogers and McClelland 2004), but instead address them at the same neural level at which perception is typically addressed.

Take, for example, the fact that when two people talk about the same thing, their eye movements can fall into remarkable synchrony (Richardson and Dale 2005). Since the two brains do not have direct neural connections, one can easily fall into the trap of describing the phenomena as a case of downward causation: thoughts and sentences (coherent high-level structures) causing neural activation (low-level elements) to generate patterns of motor output that are similar between the two people. However, human-to-human conversation can also be generally described solely at that lower level of organisation, with the invocation of higher-level structures serving as nothing more than a descriptive shortcut, not as a necessary link in the causal chain. Neural patterns in visual cortex (corresponding to scene perception) cause neural patterns in the frontal eye fields, which cause the eyes to move to a particular object in the scene, where light reflected off that object alters the neural patterns in visual cortex (corresponding to object recognition), which cause neural patterns in frontal cortex (corresponding to a concept or idea), which cause neural patterns in the left inferior frontal gyrus (corresponding to a sentence), which cause the mouth and tongue to move in complex ways, which cause the air coming out to have specific patterns of sound waves as it enters the listener's ears, which cause neural patterns in the listener's auditory cortex, which cause neural patterns in Wernicke's area (corresponding to word recognition), which cause neural patterns in frontal cortex (corresponding to a version of the transmitted idea), which cause neural patterns in visual cortex (corresponding to attention towards the relevant objects in the scene), which cause neural patterns in the frontal eye fields, which cause an eye movement to the same object that the speaker just looked at and spoke about. Thus, in this admittedly idealised lower-level description, at no point is there a higher-level cognitive structure that is non-decomposable to the lower neural and biomechanical level and which is required to connect the causal elements that produce eye-movement synchrony during conversation. However, that description is sufficiently convoluted and ungainly to make clear why many cognitive scientists would prefer to use the higher-level description of such a communication event. The point here is that there is indeed no privileged level of organisation (as noted by many authors in this issue), but of any pair of adjacent levels, it is the lower level whose causal forces generally produce the global patterns that are observed at the higher level, not the other way around. Therefore, it is always worthwhile to at least attempt to discover the underlying dynamics of any complex non-linear phenomena, even when the higher-level description is elegant and mostly successful.

2. The Game of Life

In his target article, Dale (2008) brings up John Conway's Game of Life as an example of a cellular automaton where a small number of simple underlying rules (at a lower level of description) can create very sophisticated and apparently-organised patterns of behaviour (at a higher level of description). The Game of Life consists of a two-dimensional rectangular grid of cells, each of which turns 'on' or 'off' according to two rules: (1) A cell that is off will turn on if exactly three of its eight immediate neighbours are on and (2) A cell that is on will turn off if less than two or more than three of its neighbours are on. From these two simple rules arise a series of complex patterns of activity that may appear as if there are some overarching guiding principles in effect to govern their behaviour.

For example, when a glider (a roughly triangular shape) propagates across the lattice in the Game of Life, it looks to the observer as if it is a coherent object that is intentionally wiggling its way in a particular direction, almost as if it were late for a meeting. But a glider is not an object. The cells that 'make it up' do not move. A glider is a dynamic pattern of information (e.g. Kelso 1995; Freeman 2001) whose activity is created through the application of two extremely simple and local rules. This kind of surprising emergence has a tendency to carry with it an air of mysteriousness. However, focussing solely on the fact that coherent behaviour can emerge unpredictably from local simple rules, and then treating those behaviours as if their unpredictability protects them from reductionism, runs counter to the very heart of what science is about.

There are two versions of the lesson that can be learned from Conway's game of life. The first version is that complex behaviour can emerge mysteriously in a complex system, and since this process is difficult to predict from the underlying elements, we should focus solely on laws that describe the global behaviour. This version of the lesson encourages us to ignore the constituent elements of the emergence process. It shares much with Fodor's (1974) call to view cognitive science as a 'special science', in which an understanding of the higher (psychological) level of analysis is all that we need to obtain. However, this view seems to be throwing its hands up at the complexity of the brain and too readily allowing us to content ourselves with the easier level of analysis. The second version of the lesson from Conway's Game of Life is that we must recognise that complex behaviour can emerge where there is no overarching rule, and seemingly 'guided' behaviour can emerge from a system in which there is no guidance to be found – at all times the dynamics are being driven solely by those local simple rules at the subsymbolic level. That is, rather than being comforted by meaningful things emerging out of the interaction of meaningless elements in the Game of Life, we should instead be awestruck by the powerful illusion of meaningfulness that emerges from those tiny elements that we know perfectly well carry no meaning. And we can import that lesson into our study of cognition.

A pluralist approach that allows the level of global coherent units of thought to exist without being reducible to their neural and biomechanical underpinnings, and perhaps even to exert some downward causation on those underpinnings, is exactly the spurious convergence that this special issue projects on the horizon between emergentist and symbolic perspectives in cognitive science. Even Robert Laughlin (whom Dale invokes as his physics expert on emergentism) admits in his preface that 'All physicists are reductionists at heart, myself included. I do not wish to impugn reductionism so much as to establish its proper place in the grand scheme of things'. (Laughlin 2005, p. xv). Thus, the message to receive from emergentism is not that reductionism is wrong. The global coherent phenomena that emerge in a complex non-linear system are undeniably caused by the interactions of the tiny elements that make up that system, not the other way around. The message to receive is that linear component-wise reductionism is wrong. That is, if our study of cognition were to involve formulating our intuitive guesses as to how it is carved up into subcomponents (e.g. a vision module, a language module, a working memory module, an associative memory module and a central executive module), and looking for those subcomponents at the next lower level of analysis (neural systems), then the facts surrounding emergentism will surely trip us up. However, if our study of cognition involves an integrated analysis of how coherent cognitive behaviour arises from complex non-linear interactions between neuron-like units (Rogers and McClelland 2004), between cognitive subsystems (Spivey 2007) and between action-perception synergies (Turvey 2007), along with the conceptual tools to analyse this coordination (Van Orden 2008), then emergentism will be our friend.

3. Symbolic dynamics

Such an integrated analysis is likely to receive some temporary assistance from symbolic dynamics – which Edelman (2008) invokes to support a computational equivalence between symbolic and dynamical accounts of mind. Symbolic dynamics (e.g. Crutchfield 1994; beim Graben 2004; see also, Dale and Spivey 2005) may at first appear to be a happy compromise between those theories that emphasise discrete states and those that emphasise dynamic processes. In symbolic dynamics, regions of a volumetric state space are discretely delineated into sections that correspond to different categorical symbols. As a continuous trajectory moves through this state space, upon entering one of these partitioned regions, the discrete symbol associated with that region is emitted. As the trajectory moves out of one region and into another categorical region, the new discrete region's symbol is likewise emitted, forming a sequence of categorically discrete symbols. This characterisation shares much with classical cognitive science's assumption that perception may be distributed and dynamic in its processes, but when it communicates to cognition, the result is a string of logical symbols (e.g. Hummel and Holyoak 2003; Jilk et al. 2008).

As Edelman (2008) points out, symbolic systems and dynamic systems can actually be equivalent, given certain parameters, as long as the computations being performed by each are the same. For example, in very special circumstances, the emitted string of symbols can be used to reconstruct a version of the original state-space trajectory (e.g. Takens 1981; Crutchfield 1994). However, it is the limitations of these parameters that should give us pause in considering symbolic dynamics as a viable way of resolving the debate between dynamic and symbolic systems. For example, there is the non-trivial problem of threshold setting. In creating models that use symbolic dynamics, setting the exact thresholds of the categorical regions (the partition), indicating when the trajectory has left one discrete region and entered another, is a tricky business with dire consequences if done incorrectly (Bollt, Stanford, Lai, and Zyczkowski 2001). Even with statistically sophisticated techniques, slight imprecision in these thresholds may occur, and as a non-linear trajectory travels through this state space, misleading strings of symbols may be emitted as a result. These 'ungrammatical' strings lead to dramatic misrepresentations in the receiving system regarding the sending system's actual internal dynamics.

Thus, although symbolic dynamics may seem to be classical cognitive science's last and only hope to remain relevant to the study of the mind, even it cannot save the modular rule-based approach to modelling human cognition. In symbolic dynamics, the more regions you partition and the more frequently you sample from the original continuous trajectory (in a sense, the 'closer' you are to the original signal), the more accurate your symbolic description of the system will be. Correspondingly, Atmanspacher (2000) treats symbolic dynamics like the measurement problem from physics, and thus refers to the original continuous trajectory as a sequence of ontic states (what is really there) and to the symbol string as a sequence of epistemic states (how we describe what is really there). Crucially, for symbolic dynamics to rescue classical cognitive science from antiquity, we would require optimism regarding the finding of a generating partition in the state space of the human mind, which would guarantee equivalence between that ontic trajectory and any resulting epistemic symbol string. However, the complexity and high-dimensionality of that state space almost surely preclude any such optimism. As Dale and Spivey (2005) note, 'As we approach a level of complexity that matches what is accomplished in a neural substrate, or proposed cognitive processes of multiple dimensions, the likelihood of finding generating partitions drops radically' (p. 334).

Therefore, instead of relying on symbol strings to reconstruct the state space of the mind, we are encouraged (like experimental physicists) to get our laboratory measurements 'closer' to the original continuous trajectory, closer to 'the ontic stream'. We suspect that more temporally continuous measures in cognitive science will not only improve on the advances of symbolic approaches to cognition but will eventually make possible a satisfactory account of how dynamically distributed neural and biomechanical processes can produce cognitive/behavioural phenomena that give the illusion of symbolic processing.

4. The ontic stream

Rather than taking sides with any author's position in this issue, we find ourselves drawn to highlight their common unexpected agreement (on pluralism), and to disagree with that. But identifying the appropriate locus of the debate is important. A red herring has perhaps arisen due to a perspective within the dynamical systems movement in cognitive science that portrays the mind as 'non-computational' (e.g. van Gelder 1998) in a rather indiscriminate use of the term. Edelman (2008) argues convincingly, we think, that this cannot possibly be the case. As we know from work in symbolic dynamics, there are certain constrained circumstances (e.g. low dimensionality, zero noise) where the analog function being performed by a continuous dynamical system and the Turing computational function manifested by its symbolic output (via a generating partition) are formally equivalent. In fact, even when the dimensionality is high and noise is present, thus rendering a true generating partition infeasible, the continuous dynamical trajectory can still be described as performing analog computation on its own.

Thus, 'computational *versus* non-computational' is not the divide along which the symbolic/dynamical debate in cognitive science should be drawn. Discrete symbolic structures are visible at a higher, more abstract level of description (with coarser temporal and spatial resolution) than trajectories in a continuous state space. As a result, the two do not have to be mutually exclusive theoretical frameworks that naturally compete for the same currency. The only reason they are treated as mutually exclusive is because each has a bad habit of carrying as a core tenet an arbitrary and unnecessary claim that the other is wrong. Rather than being wrong or right, it is simply that one of them is a closer approximation to the actual physical phenomena of human cognition than the other.[3] In fact, it is not the case that either the symbolic or dynamic perspective can simply fit more data from the same pool and thereby win the contest; they tend to apply themselves to different pools of data. For this reason, not only it is impossible for one of them to best the other in the usual way, but it is also impossible for them to co-exist as equal pluralist partners, because one of them is (non-linearly) reducible to the other.

Modelling the mind with rules and symbols has been an extremely useful intermediate step toward understanding how human cognition works, providing a first approximation at a higher level of organisation. However, it is now time for cognitive science to continue its movement closer and closer to the original ontic stream of cognition, making finer and finer approximations in our epistemic account of the mind. We must develop an understanding of how those symbol-ish structures and rule-ish transitions emerge from non-linear interactions among lower-level (neural and biomechanical) elements that make up brain, body and environment.[4] This simultaneously reductionist and emergentist agenda, for the future of cognitive science is neither pluralist nor anti-pluralist with regard to symbols and dynamics, but is perhaps best labelled a conservative emergentism.

Notes

1. Of course, we should be wary of pluralist hybrid machines if they are going to have anything like the crash record of the Marines' hybrid aircraft, the Osprey, an impressive fixed-wing tilt-rotor aircraft that has vertical take-off and landing capability and a nasty penchant for pilot-induced oscillation.
2. The risk of granting the symbolic level too much causal power to influence its own underlying subsymbolic level is that one can find oneself sliding down this slippery slope towards claiming that re-naming Pluto as a dwarf planet will alter its solar orbit, or that being Jewish actually causes the DNA to replicate itself only through the mother's family line! In many circumstances, the symbolic level of organisation (where an entity might be designated as a dwarf planet or Jewish or both) is clearly nothing more than a descriptive or cultural convenience with no causal influence on the underlying dynamics.
3. A bit like Jilk et al's. (2008) analogy to liquid water, we use an analogy to water vapour. It is not 'wrong' to call a roughly bounded collection of water droplets that are airborne at high altitude 'a cloud', but the dynamics of clouds are better understood when one takes into account the fact that they are a collections of water droplets. Indeed, models of weather prediction would be in big trouble if they treated clouds as non-decomposable entities that rigidly push one another around!
4. Indeed, Jilk et al. (2008) confess that even after their pluralist collaboration of grafting a symbolic model onto a subsymoblic model, this critical phase transition from subsybmolic to symbolic was not elucidated: 'In attempting to connect the two architectures functionally, it became clear that the question of how symbols arise is unanswered in both'.

References

Atmanspacher, H. (2000), 'Ontic and Epistemic Descriptions of Chaotic Systems', in *Computing Anticipatory Systems: CASYS 99*, ed. D. Dubois, Berlin: Springer, pp. 465–478.

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

Bollt, E., Stanford, T., Lai, Y., and Zyczkowski, K. (2001), 'What Symbolic Dynamics do we Get with a Misplaced Partition? on the Validity of Threshold Crossings Analysis of Chaotic Time-Series', *Physica D*, 154, 259–286.

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

Dale, R., and Spivey, M. (2005), 'From Apples and Oranges to Symbolic Dynamics: A Framework for Conciliating Notions of Cognitive Representation', *Journal of Experimental and Theoretical Artificial Intelligence*, 17, 317–342.

Emmeche, C., Køppe, S., and Stjernfelt, F. (2000), 'Levels, Emergence, and Three Versions of Downward Causation', in *Downward Causation*, eds. P. Andersen, C. Emmeche, N. Finnemann and P. Christiansen, Århus: Aarhus University Press, pp. 13–34.

Fodor, J.A. (1974), 'Special Sciences (or: The Disunity of Science as a Working Hypothesis)', *Synthese*, 28, 97–115.

Freeman, W.J. (2001), *How Brains make up their Minds*, New York: Columbia University Press.

Ganis, G., Thompson, W., and Kosslyn, S. (2004), 'Brain Areas Underlying Visual Mental Imagery and Visual Perception: An fMRI Study', *Cognitive Brain Research*, 20, 226–241.

Hummel, J., and Holyoak, K. (2003), 'A Symbolic-Connectionist Theory of Relational Inference and Generalization', *Psychological Review*, 110, 220–264.

Kelso, J. (1995), *Dynamic Patterns: The Self-Organization of Brain and Behavior*, Cambridge, MA: MIT Press.

Kim, J. (2000), 'Making Sense of Downward Causation', in *Downward Causation*, eds. P. Andersen, C. Emmeche, N. Finnemann and P. Christiansen, Århus: Aarhus University Press, pp. 305–321.

Laughlin, R.B. (2005), *A Different Universe*, New York: Basic Books.

Lupyan, G., and Spivey, M. (2008), 'Perceptual Processing is Facilitated by Ascribing Meaning to Novel Stimuli', *Current Biology*, 18, r410–r412.

McRae, K., de Sa, V., and Seidenberg, M. (1997), 'On the Nature and Scope of Featural Representations of Word Meaning', *Journal of Experimental Psychology: General*, 126, 99–130.

Puri, A.M., and Wojciulik, E. (2008), 'Expectation both helps and hinders object perception', *Vision Research*, 48, 589–597.

Richardson, D., and Dale, R. (2005), 'Looking to Understand: The Coupling between Speakers' and Listeners' Eye Movements and its Relationship to Discourse Comprehension', *Cognitive Science*, 29, 1045–1060.

Rogers, T., and McClelland, J. (2004), *Semantic Cognition: A Parallel Distributed Processing Approach*, Cambridge, MA: MIT Press.

Spivey, M.J. (2007), *The Continuity of Mind*, New York: Oxford University Press.

Takens, F. (1981), 'Detecting Strange Attractors in Turbulence', in *Dynamical Systems and Turbulence*, eds. D. Rand and L. Young, Berlin: Springer-Verlag, pp. 366–381.

Turvey, M.T. (2007), 'Action and Perception at the Level of Synergies', *Human Movement Science*, 26, 657–697.

van Gelder, T. (1998), 'The Dynamical Hypothesis in Cognitive Science', *Behavioral and Brain Sciences*, 21, 615–665.